

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2004, 08:44:05 ; Search time 4394 Seconds
(without alignments)
10712.457 Million cell updates/sec

Title: US-09-839-894-9

Perfect score: 1086

Sequence: 1 atgaataagattttatttat.....caagtagtcaaacactctag 1086

Scoring table:

IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hcg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vb.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1086	100.0	1086	6	AX352333 Sequence
2	1086	100.0	7239	1	AF296132 Escherich
3	1086	100.0	7239	6	AX352351 Sequence
4	1086	100.0	7239	6	AX741418 Sequence
5	852.4	78.5	7174	1	ECOCFAIA
6	310.8	28.6	5336	1	ECCOCCD
7	310.8	28.6	5336	6	AX741413 Sequence
8	309.6	28.5	6956	1	AY216491 Escherich
9	298.8	27.5	5798	1	ECCOTABCD
10	298.8	27.5	5798	6	AX741415 Sequence
11	63.4	5.8	1141	6	AX083744 Sequence
12	57.6	5.3	61864	3	C3Y50B5A
13	53.6	4.9	5230	8	AF253047
14	50.6	4.7	29143	3	AC115594
15	50.6	4.7	88549	3	AC116979 Dictyoste
16	50.2	4.6	133477	2	AC006278 Plasmodiu
17	50.2	4.6	231551	3	AE014844
18	50	4.6	2000	6	AX655393 Sequence
19	49.6	4.6	253924	3	AE014822 Plasmodiu
20	49.4	4.5	10279	6	AX279996 Sequence
21	49.4	4.5	10279	6	AX346493 Sequence
22	49.4	4.5	10279	6	AX356452 Sequence
23	49.2	4.5	23151	3	AE014842
24	48.6	4.5	1141	6	AX083744
25	48.2	4.4	9504	6	AX323793 Sequence
26	48.2	4.4	349980	6	AX344570 Sequence
27	48	4.4	3549	3	AF310884 Dictyoste
28	47.8	4.4	128683	2	AC135793 Oryza sat
29	47.6	4.4	169794	2	AC004688 Plasmodiu
30	47.6	4.4	196149	2	AC004709 Plasmodiu
31	47.6	4.4	252650	3	AE014847 Plasmodiu
32	47.2	4.3	12940	3	AE001386 Plasmodiu
33	46.8	4.3	105682	3	AC116957_3
34	46.6	4.3	110000	2	PFMAL13_09
35	46.6	4.3	110000	2	PFMAL13_10
36	46.6	4.3	132254	3	PFMAL6PL_05
37	46.6	4.3	132254	3	AC116330 Dictyoste
38	46.2	4.3	76568	3	MSREV
39	46.2	4.3	110000	3	AC116305_1
40	46	4.2	111882	3	AC115612 Dictyoste
41	46	4.2	153751	3	AC116551 Dictyoste
42	46	4.2	176436	2	EX323051 Danio rer
43	46	4.2	236521	2	AC129274 Rattus no
44	46	4.2	331039	3	AC116988 Dictyoste
45	45.6	4.2	251448	3	AE014819 Plasmodiu

ALIGNMENTS

RESULT 1
AX352333
LOCUS AX352333 1086 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 9 from Patent WO0181582.
ACCESSION AX352333
VERSION AX352333.1 GI:18617616
KEYWORDS
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.

REFERENCE

AUTHORS

TITLE

Pred. No. is the number of results predicted by chance to have a

JOURNAL Patent: WO 0181582-A 9 01-NOV-2001;
University of Maryland, Baltimore (US)
FEATURES Location/Qualifiers
source 1..1086
/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"
CDS 1..1086
/notes="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAD22854.1"
/gb_xref="GI:18617617"
/translation="MNKILFIPTFVSFLTFVAVSDKIPGDESITNIIFSPDRNES
SPXNIIINHTIYAVSESHITVDMFTFLCLSHNTLNACPTSENSSSVSGETNITL
QFTEKSLIARELIQYKOLLPKSNCPGSLTNSAHFNCKNNAAGSLVLYIPAG
ELKNLPGGLWDAILKLRVRYSEYTGTYINITIKLTDKGNLQIWLPPKSDAVID
LAURPTGGTYIGNSVDMCFDGYSTNSSLSEKRDNNPKPDGKPYLRKINDDTKE
IAYTLKLLAGKSLVPTNGSLNIAADAASLETKWNRITAVTMEIISVFLVCFPGRLQ
DAKVNPEAQYMGNIINVTTPSSQL"

Query Match 100.0%; Score 1086; DB 6; Length 1086;
Best Local Similarity 100.0%; Pred. No. 3.9e-223;
Matches 1086; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGAATAAGATTATTTATTTTACATGTTCTTCTTCAGTACTTTTACATTTGCT 60
DB 1 ATGAATAAGATTATTTATTTTACATGTTCTTCTTCAGTACTTTTACATTTGCT 60
QY 61 GTATCGGCAGATAAAATCCCGGAGATGAAGCAATACTAATATTTTGGCCCGCTGAC 120
DB 61 GTATCGGCAGATAAAATCCCGGAGATGAAGCAATACTAATATTTTGGCCCGCTGAC 120
QY 121 AGGAACGATCTCCCAACAAATATATTAATTAACCATATTAAGCATACAGTGAA 180
DB 121 AGGAACGATCTCCCAACAAATATATTAATTAACCATATTAAGCATACAGTGAA 180
QY 181 AGTCATCTCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
DB 181 AGTCATCTCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 241 GGACATGTCCAAACAGTCAGAAATCCATGACAGTTCATCGTCAGCGGTGAACAAATA 300
DB 241 GGACATGTCCAAACAGTCAGAAATCCATGACAGTTCATCGTCAGCGGTGAACAAATA 300
QY 301 ACATTACATTTACGGAAGAAAGAGTTTATATAAAGAGAGCTACAAATTAAGGCTAT 360
DB 301 ACATTACATTTACGGAAGAAAGAGTTTATATAAAGAGAGCTACAAATTAAGGCTAT 360
QY 361 AAACAAATTTGTTCAAAAGTGTAACTGCGCCATCCGCGCTTAACACTCAGCTCAT 420
DB 361 AAACAAATTTGTTCAAAAGTGTAACTGCGCCATCCGCGCTTAACACTCAGCTCAT 420
QY 421 TTTAACTGTATAAAGACCGCGCTTCAGTGCAAGTTTATATTTATATATTCCTGCTGGC 480
DB 421 TTTAACTGTATAAAGACCGCGCTTCAGTGCAAGTTTATATTTATATATTCCTGCTGGC 480
QY 481 GAACATTAATAATTTGCTTTTGGTGTATCTGGATGCTACTCTCAAGTTAAGAGTAAAA 540
DB 481 GAACATTAATAATTTGCTTTTGGTGTATCTGGATGCTACTCTCAAGTTAAGAGTAAAA 540
QY 541 AGACGATATAGTCAGACCTATGGAACCTTACACTATAAATATCACTATTAATTAATCAAT 600
DB 541 AGACGATATAGTCAGACCTATGGAACCTTACACTATAAATATCACTATTAATTAATCAAT 600
QY 601 AAGGAATATTCAGATATGTTACCTCAGTCAAAAGTGCAGTCCGGTGCATCTTAAC 660
DB 601 AAGGAATATTCAGATATGTTACCTCAGTCAAAAGTGCAGTCCGGTGCATCTTAAC 660
QY 661 TTGCTTCAACTGTGGGGGCACATATATTTGGAGAAATCTGTTGATGATGCTTTTAT 720

ORIGIN
Query Match 100.0%; Score 1086; DB 6; Length 1086;
Best Local Similarity 100.0%; Pred. No. 3.9e-223;
Matches 1086; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGAATAAGATTATTTATTTTACATGTTCTTCTTCAGTACTTTTACATTTGCT 60
DB 1 ATGAATAAGATTATTTATTTTACATGTTCTTCTTCAGTACTTTTACATTTGCT 60
QY 61 GTATCGGCAGATAAAATCCCGGAGATGAAGCAATACTAATATTTTGGCCCGCTGAC 120
DB 61 GTATCGGCAGATAAAATCCCGGAGATGAAGCAATACTAATATTTTGGCCCGCTGAC 120
QY 121 AGGAACGATCTCCCAACAAATATATTAATTAACCATATTAAGCATACAGTGAA 180
DB 121 AGGAACGATCTCCCAACAAATATATTAATTAACCATATTAAGCATACAGTGAA 180
QY 181 AGTCATCTCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
DB 181 AGTCATCTCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 241 GGACATGTCCAAACAGTCAGAAATCCATGACAGTTCATCGTCAGCGGTGAACAAATA 300
DB 241 GGACATGTCCAAACAGTCAGAAATCCATGACAGTTCATCGTCAGCGGTGAACAAATA 300
QY 301 ACATTACATTTACGGAAGAAAGAGTTTATATAAAGAGAGCTACAAATTAAGGCTAT 360
DB 301 ACATTACATTTACGGAAGAAAGAGTTTATATAAAGAGAGCTACAAATTAAGGCTAT 360
QY 361 AAACAAATTTGTTCAAAAGTGTAACTGCGCCATCCGCGCTTAACACTCAGCTCAT 420
DB 361 AAACAAATTTGTTCAAAAGTGTAACTGCGCCATCCGCGCTTAACACTCAGCTCAT 420
QY 421 TTTAACTGTATAAAGACCGCGCTTCAGTGCAAGTTTATATTTATATATTCCTGCTGGC 480
DB 421 TTTAACTGTATAAAGACCGCGCTTCAGTGCAAGTTTATATTTATATATTCCTGCTGGC 480
QY 481 GAACATTAATAATTTGCTTTTGGTGTATCTGGATGCTACTCTCAAGTTAAGAGTAAAA 540
DB 481 GAACATTAATAATTTGCTTTTGGTGTATCTGGATGCTACTCTCAAGTTAAGAGTAAAA 540
QY 541 AGACGATATAGTCAGACCTATGGAACCTTACACTATAAATATCACTATTAATTAATCAAT 600
DB 541 AGACGATATAGTCAGACCTATGGAACCTTACACTATAAATATCACTATTAATTAATCAAT 600
QY 601 AAGGAATATTCAGATATGTTACCTCAGTCAAAAGTGCAGTCCGGTGCATCTTAAC 660
DB 601 AAGGAATATTCAGATATGTTACCTCAGTCAAAAGTGCAGTCCGGTGCATCTTAAC 660
QY 661 TTGCTTCAACTGTGGGGGCACATATATTTGGAGAAATCTGTTGATGATGCTTTTAT 720

Db 661 TTGCTTCAACTGTGGGGGCACATATATTTGGAGAAATCTGTTGATATGCTTTTAT 720
QY 721 GATGGATATAGTACTAACAGCAGCTCTTTGGAGATAAGATTTTCAGGATAACAATCCTAAA 780
DB 721 GATGGATATAGTACTAACAGCAGCTCTTTGGAGATAAGATTTTCAGGATAACAATCCTAAA 780
QY 781 TCTGATGGGAAATTTTATCTAAGGAAAAATTAATGATGACACCAAGAAATTCATATACT 840
DB 781 TCTGATGGGAAATTTTATCTAAGGAAAAATTAATGATGACACCAAGAAATTCATATACT 840
QY 841 TTGCTACTCTCTTGGCGGTAAAAAGTTTAACTCCCAACAAATGGAACCTCATTAATAATT 900
DB 841 TTGCTACTCTCTTGGCGGTAAAAAGTTTAACTCCCAACAAATGGAACCTCATTAATAATT 900
QY 901 GCTGACGACGCTTCTCTGGAACAAACTGGAATAGAAATACAGCTGTACACATGCCAGAA 960
DB 901 GCTGACGACGCTTCTCTGGAACAAACTGGAATAGAAATACAGCTGTACACATGCCAGAA 960
QY 961 ATCAGTGTTCGGTGTGTGTTGGCTGGACGTTTGAATTTGCAATTTGCAAAAGTGGAAAT 1020
DB 961 ATCAGTGTTCGGTGTGTGTTGGCTGGACGTTTGAATTTGCAATTTGCAAAAGTGGAAAT 1020
QY 1021 CCGGAGCTGCAATATATGCGGTAAATTAATTAATGTTACTTTTCAACCAAGTAGTCAAAACA 1080
DB 1021 CCGGAGCTGCAATATATGCGGTAAATTAATTAATGTTACTTTTCAACCAAGTAGTCAAAACA 1080
QY 1081 CTCTAG 1086
DB 1081 CTCTAG 1086

RESULT 2
AF296132 7239 bp DNA linear BCT 27-FEB-2003
LOCUS Escherichia coli csa operon, complete sequence, and insertion
DEFINITION sequence ISI.
ACCESSION AF296132
VERSION AF296132.1 GI:15419711
KEYWORDS
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE 1. (bases 1 to 7239)
Altboum, Z., Levine, M.M., Galen, J.E. and Barry, E.M.
AUTHORS Genetic Characterization and Immunogenicity of Coli Surface Antigen
TITLE 4 from Enterotoxigenic Escherichia coli when It Is Expressed in a
Shigella Live-Vector Strain
JOURNAL Infect. Immun. 71 (3), 1352-1360 (2003)
MEDLINE 22483692
PUBMED 12595452
REFERENCE 2. (bases 1 to 7239)
Altboum, Z.D., Levine, M.M., Galen, J.E. and Barry, E.M.
AUTHORS Isolation and characterization of ETEC CS4 fimbriae encoding genes,
TITLE and their expression in Shigella flexneri 2a guaba strain CVD 1204
JOURNAL Unpublished
REFERENCE 3. (bases 1 to 7239)
Altboum, Z.D., Levine, M.M. and Barry, E.M.
AUTHORS Direct Submissior
TITLE Submitted (14-AUG-2000) Center for Vaccine Development, University
JOURNAL of Maryland, School of Medicine, 685 W. Baltimore Street,
Baltimore, MD 21201, USA
FEATURES Location/Qualifiers
1..7239
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="E1181A"
/serotype="O25:H42"
/db_xref="taxon:562"
/note="enterotoxigenic"
283..6849
/gene="csa operon"
/note="encodes ETEC CS4 fimbrial proteins; CFA/IV"

```

283..999
/gene="csaA"
CDS
/notes="colonization factor antigen a; may stabilize the
major pilin protein"
/codon_start=1
/transl_table=11
/product="periplasmic chaperone-like protein CsaA"
/protein_id="AAK97134.1"
/db_xref="GI:15419712"
/translation="MEKFLCCLSLITPPVANFMFLYPIISKDKNGNSLISVYSKS
KEIQIKITKTIINGCTEEHEVDNMDGGVFWTPQKVLPAQGSKSLRTPQRP
KKEIYRVFVAVPKSKNVINKKLTBELSVNLIYAAIRSLPSEQNLISLRNA
KMLIIYNGNVRAGVKDIYFCKSSNIDDSVCRKTKNKIYPERKSFDTLVNNNF-SYVF
IKLHEDIEKEQSLTLQKV2"
1028..1531
/gene="csaB"
CDS
1028..1531
/gene="csaB"
/notes="colonization factor antigen b"
/codon_start=1
/transl_table=11
/product="CSA major fimbriae subunit CsaB"
/protein_id="AAK97135.1"
/db_xref="GI:15419713"
/translation="MKLKITIGAMALTTFMVAMSASAVEKNIITVTASVPTDIILOAD
GSSLPATVELTYSFAASRPENYKIATKVNTVINKVLKLVNDPKLTNVLDTKQLP
ITVSYGKILSTADVTPEPALNFGTSGVTVGVSSQDLVIGATTAAQAPTAGNYGVVS
IMTLAS"
1589..4192
/gene="csaC"
CDS
1589..4192
/gene="csaC"
/notes="putative outer membrane protein involved in the
transport of the major fimbriae subunits across the outer
membrane; colonization factor antigen c"
/codon_start=1
/transl_table=11
/product="usher protein CsaC"
/protein_id="AAK97136.1"
/db_xref="GI:15419714"
/translation="MPKKNTLITITIAMLTPIYSVFGSDIPIENSPDLNGRQDEYEVKL
YQGLGHLIKITPTPHIKYSPESLIDKINVKEKKKLSVLFTPSFNGMSCOGN
ATIQNCNKITKSDVPIVDDVNVNLFIGNFELDEAHNDYHQLSRVKKAFQS
QTINVSQKYSLSVSGNSALGITDTSYAVLNMWNYNKFNGYSNNEITINSLEPRH
DLDKRYYQFGRDRLDLSQISGNFENLLPLPDIDGRTGTQSYIKNTDKFIASP
VTMLTFNRVFAEPRNNQLLVYLDGYNELDPTARLPYGSYDLKLIKENTQKFAEP
IIPFNKGRSIIQMDQWDFLOGGNIINDKRVIEKNNHKS VNAGLRPIITKNISVO
QGSVIDNKYVGSILKNSGILLSGLSNPFSLMGDNAGKNYSISYTDGFSLSFYH
NKRVDNCRNINAGSGCTESIASLSPLIGWTILGISDYSESIVKHHLSBYG
FYNNLYKRTORWQITSTSLKMDYNFMPAIGYNSPQRLTDKGGYISVLTPLAS
RENSNAGYSYNSRGKYSNBLFVDGYMTSTNGDYHEVRMFNKRHNHAGRLSGR
INNRFDGLNGSFMKNRNTNSHSLTGTYSNFALTSDGFWYGGSGASGLTKLAGCI
IKVKSNDIKNIVKVTGALYGYDLSGSDNAPLPVPALTPASLIIDBNNGDKNISVL
APTNMDFILPGNVFVEIETKVSYSYIGGDKNGTPLSGAHVNEPWHVILDEGGF
SREYCNKTEILFKERTYTQCLGKNKVKHGLVFGVDVCDVNSTSLPDEFVKNR
VDLLAKNDKG"
4196..5281
/gene="csaE"
CDS
4196..5281
/gene="csaE"
/notes="probably found at the fimbriae tip and involved in
attachment of the bacteri; colonization factor antigen e"
/codon_start=1
/transl_table=11
/product="minor pilin protein CsaE"
/protein_id="AAK97137.1"
/db_xref="GI:15419715"
/translation="MNKILFIPTFSSVLTFAVSADKIPGDESITINFGPRDRNES
SPKHNLNHIIYVSSHTLYDRMTPLCLSSHTNGACPTGENPSSSVSGETITL
QTEKSLIKRQLQKGYKOLLFKSVNCPGLTSAHFNCKNAASGASLYLYPAG
ELKNLPFGGIWDATLKRVRKRYSETYGYTITIKLTDKGNIQIWLQPKFSARVD

```

```

INLPTGGGTYIGRNVDMCFYDGYSTNSSLSIRIQDNNPKSDGKLYLRKINDDTKE
IAYTLISLAGSLATPTNGTSLNIADAASLETNWRITAVTNPISVPLWCPERLQL
DAKVENPEAQYMGNIINVTFTPSSQTL"
complement(5790..6119)
/gene="csaD"
/notes="truncated; CsaD"
6096..6870
/repeat_region
6143..6427
/insertion_seq="IS1"
/gene="islA"
/notes="truncated; Insa"
6346..6849
/gene="islB"
6346..6849
/gene="islB"
/codon_start=1
/transl_table=11
/product="InsaB"
/protein_id="AAK97138.1"
/db_xref="GI:15419716"
/translation="MPGNSPHYGRWPFQHDFTSLKLRPOSVTSRIQPSDVIYCAEMD
EQWYGVGAKSRQRFWLFYADRLEKTVVAHVFGERTMATLRLMSLLSPFDVIMTDC
WPLYESRLKGLHVISKRYTQRTERHNLNLRQHLARLGRKLSLSPKSVLHDKVIGHY
LNKHYQ"
ORIGIN
Query Match 100.0%; Score 1986; DB 1; Length 7239;
Best Local Similarity 100.0%; Pred. No. 3e-223;
Matches 1096; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATCAATAAGATTTTATTATTATTTTACATGTTTTTCTCTTCTCAGTACTTTTACATTGCT 60
DB 4196 ATCAATAAGATTTTATTATTATTTTACATGTTTTTCTCTTCTCAGTACTTTTACATTGCT 4255
QY 61 GTATCGGCAGATAAAATTCGCGAGATGAAAGCATAACTAAATATTTTGGCCCGCGTGAC 120
DB 4256 GTATCGGCAGATAAAATTCGCGAGATGAAAGCATAACTAAATATTTTGGCCCGCGTGAC 4315
QY 121 AGGAACGAATCTTCCGCCAAACATAATATTAATAAATTAACCATATTACAGCATACAGTGAA 180
DB 4316 AGGAACGAATCTTCCGCCAAACATAATATTAATAAATTAACCATATTACAGCATACAGTGAA 4375
QY 181 AGTCATACCTGTATCATAGGATGACTTTTTATGTTGTTCTTCTCACAATACACTTAAT 240
DB 4376 AGTCATACCTGTATCATAGGATGACTTTTTATGTTGTTCTTCTCACAATACACTTAAT 4435
QY 241 GGAGCATGTCCCAACCACTGAGATCCTAGCAGTTCATCGGTGAGCGTGAACAATAATA 300
DB 4436 GGAGCATGTCCCAACCACTGAGATCCTAGCAGTTCATCGGTGAGCGTGAACAATAATA 4495
QY 301 ACATTACAATTTACGGAAAAAGAGCTTTTAATAAAAAAGAGAGCTACAAATTTAAAGGCTAT 360
DB 4496 ACATTACAATTTACGGAAAAAGAGCTTTTAATAAAAAAGAGAGCTACAAATTTAAAGGCTAT 4555
QY 361 AAACATTAATCTGTAAAAGTGTACTGCCATCCGCGCTACACTTACCTACCTCAGCTCAT 420
DB 4556 AAACATTAATCTGTAAAAGTGTACTGCCATCCGCGCTACACTTACCTCAGCTCAT 4615
QY 421 TTTAACGTGTAATAAAAAACGGCGCTTCAGGTGGAAGTTTATTTATATATATTCCTGCTGGC 480
DB 4616 TTTAACGTGTAATAAAAAACGGCGCTTCAGGTGGAAGTTTATTTATATATTCCTGCTGGC 4675
QY 481 GAACATAAAAAATTTGCTTTTGGTGGTATCTGGGATGCTACTCTGAAGTTAAGAGTAAAA 540
DB 4676 GAACATAAAAAATTTGCTTTTGGTGGTATCTGGGATGCTACTCTGAAGTTAAGAGTAAAA 4735
QY 541 AGACGATATAGTGAGACCTATGGAACCTTACACTATAAATATACACTTAAAATTAACGTAT 600
DB 4736 AGACGATATAGTGAGACCTATGGAACCTTACACTATAAATATACACTTAAAATTAACGTAT 4795
QY 601 AAGGGAATATTCAGATATGGTTACCTCAGTTTCAAAAGTCAGCTCGCTCGCTCCTTAAC 660
DB 4796 AAGGGAATATTCAGATATGGTTACCTCAGTTTCAAAAGTCAGCTCGCTCGCTCCTTAAC 4855

```

```
QY 661 TTGCGTCCAACTGTGGGGCACATATATTTGGAAGAAATCTGTGTGATATGTGCTTTAT 720
Db 4856 TTGCGTCCAACTGTGGGGCACATATATTTGGAAGAAATCTGTGTGATATGTGCTTTAT 4915
QY 721 GATGGATATAGTACTAACACAGCTCTTGGAGATAAGATTTCAGGATAACAACTCTAAA 780
Db 4916 GATGGATATAGTACTAACACAGCTCTTGGAGATAAGATTTCAGGATAACAACTCTAAA 4975
QY 781 TCTGATGGAAATTTTATCTAAGGAAATAAATGATGACACAAAGAAATTCGATATACT 840
Db 4976 TCTGATGGAAATTTTATCTAAGGAAATAAATGATGACACAAAGAAATTCGATATACT 5035
QY 841 TTGTCACTTCTTGGCGGGTAAAGTTTAACTCCACAAATGAAAGTCAATTAATTT 900
Db 5036 TTGTCACTTCTTGGCGGGTAAAGTTTAACTCCACAAATGAAAGTCAATTAATTT 5095
QY 901 GCTCAGCAGCTTCTCTGGAACAACTGGAATGAAATTCACGCTGCACATGCCAGAA 960
Db 5096 GCTCAGCAGCTTCTCTGGAACAACTGGAATGAAATTCACGCTGCACATGCCAGAA 5155
QY 961 ATCAGTGTTCGGGTGTGTGTGGCCCTGGAGCTTTGCAATTTGGATGCAAAAGTGGAAAT 1020
Db 5156 ATCAGTGTTCGGGTGTGTGTGGCCCTGGAGCTTTGCAATTTGGATGCAAAAGTGGAAAT 5215
QY 1021 CCCGAGGCTGACAAATATATGGTAAATTAATTAATGTTACTTTCAACCAAGTAGTCAACA 1080
Db 5216 CCCGAGGCTGACAAATATATGGTAAATTAATTAATGTTACTTTCAACCAAGTAGTCAACA 5275
QY 1081 CTCTAG 1086
Db 5276 CTCTAG 5281

RESULT 3
AX352351 LOCUS AX352351 7239 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 27 from Patent WO0181582.
ACCESSION AX352351
VERSION AX352351.1 GI:18617634
KEYWORDS
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
AUTHORS Altboun, Z., Levine, M.M. and Barry, E.M.
TITLE Isolation and characterization of the cea operon (etec-cs4 pill)
JOURNAL and methods of using same
Patent: WO 0181582-A 27 01-NOV-2001;
University of Maryland, Baltimore (US)
FEATURES
Location/Qualifiers
source 1..7239
/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"
283..999
/feature="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAD22859.1"
/db_xref="GI:18617635"
/db_xref="REMBL:CAD22859"
/translation="MHKILFLLIITPFFVANFMFIPIKSLKNGNSLIRVYSK
KEIOVTKIKKINPGTSEHVMNWDGVTPOKVLIPAGGSKSLRITQFRI
KKEIYRVYFAVDPKSKENVDKLTTLSVNIIVAALIRSPSEONISLISRNA
RNHIIYNGNVRAGVDYIFCKSSNIDSCVKTHKNITPEKSPFLVNNNSYF
IKLHNEDEKSGGLIQKVP"
1028..1531
/feature="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAD22860.1"
/db_xref="GI:18617636"
/db_xref="REMBL:CAD22860"
/translation="MTKNFLYITIIAMLTPIYSVFGSDIPNSFRDLWGEQDEYEVKL
YQGLTHIRKITTTHIKYSPESILDKNVKXKKKJLVLTNPSRNGNSCCQN
ATVQNCNTIKTSVDVIVDDVDVNVNLFIGNFELDEAHNDYHQLSRNVKAFQ
QTINVSQKSLSVSGNSALGTTDTSVAIVNMWYNKFNKNGNRTINSLYFPH
DLKRYVQFGRMDRTDLQS-SCNENFNLPLPDIDGTGTQSVIKNTDVIAPR
VTNLFNFRSVEAFRANQLLVYLDGSGNELTARLPYSGYDLKLPENTOLVRE
IIPFNKRSIGMDQVFIQGGNIINDKRYIEKQNNHKSVMAGRLFTTKNISVQ
QASVDNCKNYEGLSKWNSGILSLNBEFSLWGDNAKGYTQSYSDTGFSLSPFH
NDRKVDNCGNINAGSGCYVESASLSIPLLGWTSTLGYSDTYSSVYKHIISYH
FYNONYKGTORWLTSTSLKWDYNTFPAIGIYNSRQRLTDKGGYISVLTGRAS
RENLAAGYSVNYSRGKYSNELFVDSYNTSTNGDYHEVMRPNKRNHAEGRLSR
INRPFGLNGSPMNKRNNTNNSHSLTGCYNSPALTSDFYNGGASGLTKLAGGI
IKVSNDRKKNLVKVTGALYDGLSDNDAFIPVPAITFASLIEBNNYGDKNISVL
APTNDFILFENYVPELTKVSVSYGRGFDKNGTFLSGAHVLPNEPHVILDEDDGF
SPFYTGNEKTLFLKGRITTYTCQ-GRKNVHKHGVFVGDVICYNS"SSLDFEVKPR
VQDLAKNDKG"
4196..5281
/feature="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAD22862.1"
/db_xref="GI:18617638"
/db_xref="REMBL:CAD22862"
/translation="MHKILFLLIITPFFSVLFTFAVSADKIPGDEITNIFPRDRNS
SPKHNHNTIAYSESHTYDRMTFLCLSHNTLNGACPTSENPSSSVSGENTJL
QFTEKSLIKREIQIKGYKOLLPKSVNCPSPGLTNSAHFNKNAAGSLYLIPAG
ELNLPFGGIWDATLKLVRKRYSTGYTINITIKLTDKNIQIWLPOFKSDARYD
LNLRTPGGTYIGRNSVDMCFYDGYSTNSSLERFQDNPKSDGYFLKINDTKE
IAYTSLLLAGKSLTPINGTSLNIAADALETNWRITAVTMPBEISVPLVCPQLQL
DAKVENPEACQYMNINVTTPSSQT"
```

ORIGIN

```
Query Match 100.0%; Score 1086; DB 6; Length 7239;
Best Local Similarity 100.0%; Pred. No. 3e-223;
Matches 1086; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGATAAGATTTTATTTATTTTACATTTGTTTCTCTTCAGTACTTTTACATTTGCT 60
Db 4196 ATGATAAGATTTTATTTATTTTACATTTGTTTCTCTTCAGTACTTTTACATTTGCT 4255

QY 61 GTATCGGCAGATAAATTCGCGAGATGAAGATAACTAATATTTTGGCCCGCTGAC 120
Db 4256 GTATCGGCAGATAAATTCGCGAGATGAAGATAACTAATATTTTGGCCCGCTGAC 4315

QY 121 AGGAACGAATCTTCCCAACATATAATTAATAACCATATTACAGATACAGTAA 180
Db 4316 AGGAACGAATCTTCCCAACATATAATTAATAACCATATTACAGATACAGTAA 4375

QY 181 AGTCATCTCTGTATGATAGTACTTTTATGTTTGTCTCTCACAATACACTTAAT 240
Db 4376 AGTCATCTCTGTATGATAGTACTTTTATGTTTGTCTCTCACAATACACTTAAT 4435

QY 241 GGAGCATGTCCACCACTGAGATCCCTAGCAGTTCATCGGTGAGCGGTGAAACAAATATA 300
Db 4436 GGAGCATGTCCACCACTGAGATCCCTAGCAGTTCATCGGTGAGCGGTGAAACAAATATA 4495

QY 301 ACATTAACAATTTACGAAAAAGAAAGTTTAAATAAAGACAGCTTACAAATTAAGGCTAT 360
Db 4496 ACATTAACAATTTACGAAAAAGAAAGTTTAAATAAAGACAGCTTACAAATTAAGGCTAT 4555
```


QY	361	AAACRAATTATTGTTCAAAAGTGTTAACTGCCATCGGCGCTAACACTTAACCTCAGCTCAT	420
DB	4556	AAACAATTATTGTTCAAAAGTGTTAACTGCCATCGGCGCTAACACTTAACCTCAGCTCAT	4615
QY	421	TTTAACTGTAATAAAAAACGGCGCTTCAGGTCGAAGTTTATTTATATATATTCCTGCTGGC	480
DB	4616	TTTAACTGTAATAAAAAACGGCGCTTCAGGTCGAAGTTTATTTATATATATTCCTGCTGGC	4675
QY	481	GAACTAAAAAATTTGCTTTTGTGGTATCTGGATGCTACTCTGAAGTTAAAGTAAATA	540
DB	4676	GAACTAAAAAATTTGCTTTTGTGGTATCTGGATGCTACTCTGAAGTTAAAGTAAATA	4735
QY	541	AGACGATATAGTACAGACCTATGAACTTACACTTAAATATCACTATTAAATTAACCTGAT	600
DB	4736	AGACGATATAGTACAGACCTATGAACTTACACTTAAATATCACTATTAAATTAACCTGAT	4795
QY	601	AAGGGAATATTTTACAGATATGTTTACCTCAGTTCAAAGTCAAGCTCGCGTCGATCTTAAC	660
DB	4796	AAGGGAATATTTTACAGATATGTTTACCTCAGTTCAAAGTCAAGCTCGCGTCGATCTTAAC	4855
QY	661	TTGCGTCCAACTGCTGGGGCACAATATTTGGAAGAAATTTCTGTTGATATGCTCTTTAT	720
DB	4856	TTGCGTCCAACTGCTGGGGCACAATATTTGGAAGAAATTTCTGTTGATATGCTCTTTAT	4915
QY	721	GATGGATATAGTAAACAGACCTCTTTGGAGATAAGATTTTCAGGATACAAATCTTAA	780
DB	4916	GATGGATATAGTAAACAGACCTCTTTGGAGATAAGATTTTCAGGATACAAATCTTAA	4975
QY	781	TCTGATGGAAATTTTATCTAAGGAAATTAATGATGACACCAAGAAATTTGATATACT	840
DB	4976	TCTGATGGAAATTTTATCTAAGGAAATTAATGATGACACCAAGAAATTTGATATACT	5035
QY	841	TTGTCATCTCTTGTGGCGGTAAGAAATTTAACTCCAAATGGAACGTCATTAATAT	900
DB	5036	TTGTCATCTCTTGTGGCGGTAAGAAATTTAACTCCAAATGGAACGTCATTAATAT	5095
QY	901	GCTGACGACCTCTCTGGAACAACTGGAATGATGATACAGCTGCACCTGCCAGAA	960
DB	5096	GCTGACGACCTCTCTGGAACAACTGGAATGATGATACAGCTGCACCTGCCAGAA	5155
QY	961	ATCAGTGTTCGGGTGTTGTTGGCTTGGAGTTCGAAATGGATGCAAAAGTGGAAAT	1020
DB	5156	ATCAGTGTTCGGGTGTTGTTGGCTTGGAGTTCGAAATGGATGCAAAAGTGGAAAT	5215
QY	1021	CCGAGCTGGACAAATATATGCTTAATTAATGTTACTTTCACACCAAGTACTCAACA	1080
DB	5216	CCGAGCTGGACAAATATATGCTTAATTAATGTTACTTTCACACCAAGTACTCAACA	5275
QY	1081	CTCTAG 1086	
DB	5276	CTCTAG 5281	
RESULT 4			
AX741418			
LOCUS	AX741418	7239 bp	DNA linear PAT 29-MAY-2003
DEFINITION	Sequence 8 from Patent WO03022306.		
ACCESSION	AX741418		
VERSION	AX74-418.1	GI:30524215	
KEYWORDS	Escherichia coli		
SOURCE	Escherichia coli		
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.		
REFERENCE	1		
AUTHORS	Turner, A.K., Greenwood, J., Stephens, J.C., Beavis, J.C. and Darsley, M.J.		
TITLE	Bacterial vaccine		
JOURNAL	Patent: WO 03022306-A 8 20-MAR-2003;		
FEATURES	Acambis Research Limited (GB)		
source	Location/Qualifiers		
	1..7239		
	/organism="Escherichia coli"		

/mol_type="genomic DNA"	
/db_xref="taxon:562"	
283..999	gene
/gene="csaA"	
283..999	CDS
/gene="csaA"	
/codon_start=1	
/transl_table=11	
/protein_id="CAD90914.1"	
/db_xref="GI:30524216"	
/db_xref="REMBL:CAD90914"	
/translation="MHKLPCLLSLTPFVANANFMIYPISKDLKNGSELISVYSKS	
KSIQIVKIYTKKILNPCTTEHEVDPNDGQVFTPOKVLIPAGSKSRLTQFIP	
KKEEIRVFEAVKDSKENVINDKKLTTELSNIIYAAIRLSLPEQKISLISRNA	
RKNLIYNNGNRVAGVDIYFCKSSNIDDSCKVKTKHKNIIYPEKSPDTLVNNPFSVF	
IKNFEDIRKQQLKVP"	
1028..1531	gene
/gene="csaB"	
1028..1531	CDS
/gene="csaB"	
/codon_start=1	
/transl_table=11	
/protein_id="CAD90915.1"	
/db_xref="GI:30524217"	
/db_xref="REMBL:CAD90915"	
/translation="MKLKTIGAMALTTFVAMSASAVEKNITVTASVDEPTDILQAD	
GSLPTAVELTYSPAASRENYKIATKVININVKNLVILKLVNPKLTINVLDSKQLP	
ITVSGKTKLTADVTTFEALNFGTSGVTGVSQQDLVIGATTAQAPTAGNYSVVS	
ILMTLAS"	
1589..4192	gene
/gene="csaC"	
1589..4192	CDS
/gene="csaC"	
/codon_start=1	
/transl_table=11	
/protein_id="CAD90916.1"	
/db_xref="GI:30524218"	
/db_xref="REMBL:CAD90916"	
/translation="MTKNTLYITITAMLTPTVSVSGDIPNSFRDLWGODEFYEVLK	
YQGTGLHRIKTPHIFKYPESILDKINVKKEKLSVLTNPSRNGNSCOON	
ATQVNCNIVKTSVDVIVDDVNVNVLFIGNEFLDSEAHNDYHQLSRNKKAFIQS	
QIYNVDSKYSLSVSGNSALGIDTTSVAVLNWNVKNKNGYNNERTNSLYPRH	
DLDKYYVQFGRMDRTDLSQISGNFENLLPLPDIDGIRGTGTQSYIENTDKFIASE	
VIVMLNFRSGEAFRNQLWGLDVGWNEJCTARLPYSGYDEKUKIKENTOLVREE	
IIPKRRSGIGEMQWDFIQGNTINDKRYIEKONHEKSSVNAIGRLPITKNI SVQ	
QASVIDNKNYYEGSLKWNLSGLNSLSEFSLMGONAKNGYQSYTDGSLSFYH	
NKRVDNCRNYNAGWGCYVSASLIPLLGWTSLGYSVTYSVVKHLSFYH	
FYNQIVKGTORTWOLTSSTLKMVDYNMPLAIGIYNSFORQLTDAGGYISVTLTRAS	
RENSLNAGYSYNSRKYSSNELFVDGYMTSTNNGYHEVYRPNKRNAGELSGR	
INNRFGLNGSPMKNRNTNSNHSITGGYNSFALTSCGFYWGSSAGLTKLGGI	
IKVKNDRTKNLVKVTGALYDYSLSGSDNAFIPVPAITPASLIIISNNYGDKNISVL	
APTNNDMFLLPGNVPEIETKVSYSYIGKFDKNGTPLSGAHLNEPHVLDGDF	
SFEYTGNEKTLFLKGRITTYTQLGKKNVHKGIYFVGVGVICDVSNTSSLPDEFVKPR	
VODLLAKNDKG"	
4196..5281	gene
/gene="csaE"	
4196..5281	CDS
/gene="csaE"	
/codon_start=1	
/transl_table=11	
/protein_id="CAD90917.1"	
/db_xref="GI:30524219"	
/db_xref="REMBL:CAD90917"	
/translation="MNKILIFTLFFSSVYFTFAVSADKIPGDESITNIFGPRDRNES	
SPKHNILNNHITAYSESHTYLDRMTPLCSHNTLNACPTSENPSSSVSQETNITL	
QRTXESLIRKELQIKGYKQLLPKSVNCPGLTNSAHFNCKNKNASAGSLIYI PAG	
ELKNLPFGGIMDATIKLRVKRYSYFYGTINITIKLTDKGNIQIWLQFQSDAARD	
LNLRPTGGGYIGRNSVCFYDGYSTNSSSLEIRFQDNPKSDGKFIKNNDDTKE	
IATYTSLLIAGKSLPTNGTSLNIADAASLEINWNITATWPEISVPLVCPWFLQOL	
DAKVENPEAGQYMGNIINVTTFSSQL"	
6346..6849	
/note="unnamed protein product; IS1 sequence"	
/codon_start=1	

source

```

/translation="MEGNSPHYRQWQHDFTSLKKLRPOSVTSRIOPSGSDVIVCBEMD
EQWGYVGAQRQWLFYADRLKTVVAHVFGRTWATLRLMLSLSPDVIWMTDG
WPLYESRLKGLHVISRYVTQRIERHNLRLQRLARLGRKSLSPFSKVELHDKVIGHY
LNKHVQ"
ORIGIN
Query Match 100.0%; Score 1086; DB 6; Length 7239;
Best Local Similarity 100.0%; Pred. No. 3e-223;
Matches 1086; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGAATAAGATTTTATTATTATTATTACATGTTTTCCTTCAGTACTTTTACATTTGCT 60
Db ATGAATNAGATTTTATTATTATTATTACATGTTTTCCTTCAGTACTTTTACATTTGCT 4255
QY 61 GTATCGGCAGATAAAATCCCGGAGATGAAGCATAACTAATATTTTGGCCCGGTGAC 120
Db GTATCGGCAGATAAAATCCCGGAGATGAAGCATAACTAATATTTTGGCCCGGTGAC 4315
QY 121 AGGACCAATCTCCCAACACATAATATTAATTAACCATATTACAGCATACAGTGA 180
Db AGGACCAATCTCCCAACACATAATATTAATTAACCATATTACAGCATACAGTGA 4375
QY 181 AGTCATACTCTGATGATAGATGACATTTTATGTTTGTCTTCTCAATACACTTAAT 240
Db AGTCATACTCTGATGATAGATGACATTTTATGTTTGTCTTCTCAATACACTTAAT 4435
QY 241 GGAGCATGTCAACCATGTGAAATCCCTAGCAGTTCATCGGTGAGGTGAACCAATATA 300
Db GGAGCATGTCAACCATGTGAAATCCCTAGCAGTTCATCGGTGAGGTGAACCAATATA 4495
QY 301 ACATTAATTTACGGAARAGAGTTTATATAAAGAGAGCTACAAATTAAGGCTAT 360
Db ACATTAATTTACGGAARAGAGTTTATATAAAGAGAGCTACAAATTAAGGCTAT 4555
QY 481 GAATCAAAAATTTGCCCTTTTGGTGTATCTGGGATGCTACTCTGAAGTTAAGAGTAAA 540
Db GAATCAAAAATTTGCCCTTTTGGTGTATCTGGGATGCTACTCTGAAGTTAAGAGTAAA 4735
QY 541 AGAGCATATAGTACATGAGACCTATGGAACCTTACACTATAAATACACTTAAATTAAC 600
Db AGAGCATATAGTACATGAGACCTATGGAACCTTACACTATAAATACACTTAAATTAAC 4795
QY 601 AAGGCAATATTCAGATATGTTTACCTCAGTTCAAAAGTACGCTCGCTGCATCTTAAC 660
Db AAGGCAATATTCAGATATGTTTACCTCAGTTCAAAAGTACGCTCGCTGCATCTTAAC 4855
QY 661 TTGGTCAACTGCTGGGGGCACATATATTGGAAGAAATTTCTGTGATGCTTTTAT 720
Db TTGGTCAACTGCTGGGGGCACATATATTGGAAGAAATTTCTGTGATGCTTTTAT 4915
QY 721 GATGATATAGTACTAAACAGACGCTCTTTGGAGATAAGATTTTCAGATAACAATCTAAA 780
Db GATGATATAGTACTAAACAGACGCTCTTTGGAGATAAGATTTTCAGATAACAATCTAAA 4975
QY 781 TCTGATGGAATTTTATCTAAGGAATAATATGATGACCAAGAAATTCATATACT 840
Db TCTGATGGAATTTTATCTAAGGAATAATATGATGACCAAGAAATTCATATACT 5035
QY 841 TTGTCACATCTCTTGGCGGTAAAGTTTAACTCCAAAGATGGAACGCTCATTAATAT 900
Db TTGTCACATCTCTTGGCGGTAAAGTTTAACTCCAAAGATGGAACGCTCATTAATAT 5095
QY 5036 TTGTCACATCTCTTGGCGGTAAAGTTTAACTCCAAAGATGGAACGCTCATTAATAT 5095
```

```

QY 901 GGTGACGACGCTTCTCTGAAACAAACTGGAATAGAAATTACAGCTGTCCATGCCAGAA 960
Db GGTGACGACGCTTCTCTGAAACAAACTGGAATAGAAATTACAGCTGTCCATGCCAGAA 5155
QY 961 ATCAGTGTTCGGGTGTGTGTGGCTGACAGCTTTCGAATTGGATGCAAAAGTGGAAAT 1020
Db ATCAGTGTTCGGGTGTGTGTGGCTGACAGCTTTCGAATTGGATGCAAAAGTGGAAAT 5215
QY 1021 CCCGAGCTGGCAATATATATGGTAAATTAATGTTTACTTTTCAACCAAGTACTCAACA 1080
Db CCCGAGCTGGCAATATATATGGTAAATTAATGTTTACTTTTCAACCAAGTACTCAACA 5275
QY 1081 CTCTAG 1086
Db CTCTAG 5281
RESULT 5
ECOCFAIA 7174 bp DNA linear BCT 06-VAR-1995
LOCUS
DEFINITION
E.coli CFA/I fimbrial operon: cfa a, cfa b, cfa c, cfa e, cfa d,
genes, complete cds.
ACCESSION
M55661.1 GI:145507
VERSION
cfa a gene; cfa b gene; cfa c gene; cfa d' gene; cfa e gene;
KEYWORDS
colonisation factor antigen I; regulatory protein.
SOURCE
Escherichia coli
ORGANISM
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
1 (bases 1 to 7174)
AUTHORS
Hamers, A.M., Pel, H.J., Willshaw, G.A., Kusters, J.G., van der
Zeijst, B.A., and Gastra, W.
TITLE
The nucleotide sequence of the first two genes of the CFA/I
fimbrial operon of human enterotoxigenic Escherichia coli
JOURNAL
Microb. Pathog. 6 (4), 297-309 (1989)
MEDLINE
89330163
PUBMED
2569152
REFERENCE
2 (sites)
AUTHORS
Gastra, W., Jordi, B.J.A.M., Mul, E.M.A., Hamers, A.M.,
McConnell, M.M., Willshaw, G.A., Smith, H.R., and van der Zeijst, B.A.M.
TITLE
A silent regulatory gene cfaI on region 1 of the CFA/I plasmid NTP
113 of enterotoxigenic Escherichia coli
JOURNAL
Unpublished (1990)
REFERENCE
3 (sites)
AUTHORS
Jordi, B.J., Willshaw, G.A., van der Zeijst, B.A., and Gastra, W.
TITLE
The complete nucleotide sequence of region 1 of the CFA/I fimbrial
operon of human enterotoxigenic Escherichia coli
JOURNAL
DNA Seq. 2 (4), 257-263 (1992)
MEDLINE
92329981
PUBMED
1352712
COMMENT
Original
individual source text: Escherichia coli (clone NTP 513,
isolate E7473) enterotoxigenic DNA.
FEATURES
Location/Qualifiers
source
1..7174
/organism="Escherichia coli"
/mol_type="genomic DNA"
/isolate="E7473"
/db_xref="taxon:562"
/clone="NTP 513"
/dev_stage="enterotoxigenic"
848..1564
/gene="cfa a"
848..1564
/gene="cfa a"
/codon_start=1
/trans_table=11
/product="colonisation factor antigen a"
/protein_id="AAC414.1"
/db_xref="GI:145508"
/translation="MHKLFYLLSLMAPPVANFMIPYTSKDLKNGNSBLRVYSKS
KEIQYIKYIKNIINPCTTSEYKVDIPNWDGGLNVTPQKILPAGASKSIRLTQFKIP
```

```

KKEEVYVYFEAVKPDSENVNDKKLTTELVSNIIVAAALISLPSEQNIISLNSNA
KNIIIIYNNGNVRAGVKDIYFCKSSNLDNCVKANKNIYFEKSPDTIIVNNFYSYV
IKLNHEGIEKQGLIQLKVP"
848. .904
/genes="cfa a"
905. .1561
/genes="cfa a"
/products="colonisation factor antigen a"
1593. .2105
/genes="cfa b"
1593. .2105
/genes="cfa b"
/codon_start=1
/transl_table=11
/product="colonisation factor antigen b"
/protein_id="AAC41415.1"
/db_xref="GI:145509"
/translations="MKPKTIGAMALTTFVAVSASAVEKNIIVTASVDPADLLQAD
GNALPSAVKLAVSPASKTPESRVMTQVHTNDATKKVIKVLADTPQLTDLVNSTVQMP
ISVSGGQVLSITTAKEFEAAALGYSAGVNGVSSQELVISAAPKTAGTAPTAGNYSG
VSLNVNLTGS"
1593. .1661
/genes="cfa b"
1662. .2102
/genes="cfa b"
/product="colonisation factor antigen b"
2170. .4779
/genes="cfa c"
2170. .4779
/genes="cfa c"
/codon_start=1
/transl_table=11
/product="colonisation factor antigen c"
/protein_id="AAC41416.1"
/db_xref="GI:145510"
/translations="MKKKKRVAVALSVALIPYIGVTGDI PDSFRLWGEODEPFEV
KLXGOTLGERIKTTHTKFYSPESILDKINLKKKEKELSVFFTNSPSRNGMSQ
GNTTQYCNQYIKTKVDVLDVDMVNC.FIGNEFLDSEAHNDEHQLSRVVKAFI
QSQTLNLSGKRYKLSISGNSALGIDTYSYAVLNMMNKNKSGNSNEKTINSYIF
RHLDKRYQYQFMRDITLQSISGFNFNLLPLPIDIGRTGTTQSYIKNTDKFIA
SPVTMLNPSRFSENRDOLLGWYLDGWNELDTARIPYGSYDLKLI FENTQJVR
E3IIPNKRGSIGDMQWDIFVQGNIVNDNDYIEKNNHKSINTEGRALPTKNI
VOQGSVIDNKSIVYEGSLKNSILSGLSNSESFLWGNKAKNYOSISYTDGFSLF
YHNDKVDNCGRNYNAGWCYSYSASUSIPLLGWITSLIGSDTISVESYKSHLSE
YGFYQNIYKGRQWLTSYSLKMDXNEMPAIGIYNSQQLDKGQYISVITLR
ASRNLGYSYNGRYNSNEFLVDGYMTSTNRGDYHEAGMRNKNRHNABGRLS
GRINRFGLINGSFSMKNRNTSNHSLTGCYNSFALTS DGFYWGGSFAGLTLAG
GIIVKSNDDTKNIVVTGTLGYDSI GSNNDNAPI PVPAITPASLIIEDNNYGDNIS
ILAPTNNMPLPGNVYVPEIETKVSYSVIGRGFDNCTPLSGAHVLPNPHVLLDRG
GRSFYTGNEKTLFLKGRTIYTCQLGKKNVHKGI VFGDVICDINSSTSLPDEFVN
PRVQDLLAKNDG"
2170. .2235
/genes="cfa c"
2236. .4776
/genes="cfa c"
/products="colonisation factor antigen c"
4783. .5865
/genes="cfa e"
4783. .5865
/genes="cfa e"
/notes="unpublished sequence"
/codon_start=1
/transl_table=11
/product="colonisation factor antigen e"
/protein_id="AAC41417.1"
/db_xref="GI:145511"
/translations="MKKILFIPTLFRSSGEFTFAVSADKNPGENMNTNIGPHDRGS
SPYINILNLTYSRHHLYDRMSFICLSSQNTLNGACPSDAPGTAIDGENTIL
QPTKRSLLIKRELQIKYQKFLFNKANCPSKALNSHFCQNEEQSGATLSYIPAG
ELNKLFPFGVYNAVLKLNKRRDYDTYGTITINVLITDKNIGIWLQIPKSNARVD
LNLRPTGGGYTIGRNSVDMCFYDGSYTNSSSLIRIFODDNSKSDGKPYLKKIKNDSDKE
LVYTLISLLAGKNIPTNGQALNINTASLETNNKRIITAVTMEPISVPVLCWPGRLQLD
AKVNPEAGQYMGNIKIITFPSSQTL"

```

```

sig_peptide 4783. .4842
mat_peptide /gene="cfa e"
4843. .5862
/gene="cfa e"
/product="colonisation factor antigen e"
complement(6388. .6822)
/gene="cfa d"
complement(6388. .6822)
/gene="cfa d"
/notes="truncated CfaD-like protein with a stop codon and
frameshift deletions between bp 6029 and 6824."
/codon_start=1
/transl_table=11
/product="colonisation factor antigen d"
/protein_id="AAC41418.1"
/db_xref="GI:145512"
/translations="MDPKYTEREKIMKINNMIHIYVLTSGNCIMDIYSEEEKIFCS
SNRLYFLEGVNISVRIQKILSEKPYIAFLNGDILRLKNAIIVGMSKIDINDC
RNMSRKIIKTEVDKLLDLVKININSYDDSVFSSLIYLIISKI"
ORIGIN
Query Match 78.5%; Score 852.4; DB 1; Length 7174;
Best Local Similarity 87.2%; Pred. No. 4.7e-173;
Matches 947; Conservative 0; Mismatches 136; Indels 3; Gaps 1;
QY 1 ATCAATAAGATTTTATTTTATTTTACATTGTTTCTCTCTCAGTACTTTTACATTGCT 60
DB 4783 ATGAATAAAATTTTATTTTATTTTACATTGTTTCTCTCTCAGTACTTTTACATTGCT 4842
QY 61 GTATCGGCAGATAAAATTCGCGAGATGAAAGCATAACTAATATTTTGGCCGCGCTGAC 120
DB 4843 GTATCGGCAGATAAAATTCGCGAGATGAAAGCATAACTAATATTTTGGCCGCGCTGAC 4902
QY 121 AGGAACCAATCTTCCCCCAACATAATATTAATAAACAACATATTACAGATACAGTCAA 180
DB 4903 AGCGCGGATCTTCCCCCATATATAATCTTAAATCTTATCTTACAGCATACAATGGA 4962
QY 181 AGTCATACCTCTGTATGATAGTAGTACTTTTATGTTTCTCTCTCACAATACACTTAAT 240
DB 4963 AGCCATCATCTGTATGATAGTAGTACTTTTATGTTTCTCTCTCACAATACACTTAAT 5022
QY 241 GGAGCATGTCACACCACTGAGAACTCTAGCAGTTCATCGGTGAGCGGTGAAAACAAATATA 300
DB 5023 GGAGCATGCCAAGCAGTGAAGCCCTGGCAGCTCTACAATGATGGCGAAAACAATATA 5082
QY 301 ACATTACAAATTTACGAAAAAAGAAATTAATAAATAAAGAGAGCTACAAAATAAGGCTAT 360
DB 5083 ACATTACAAATTTACGAAAAAAGAAATTAATAAATAAAGAGAGAACTGCAAAATTAAGGCTAT 5142
QY 361 AACCAATTTTGTTCAAAAGTGTACTCGCCCTCCGCTACACCTAACCTAACCTCAGCTCAT 420
DB 5143 AACCAATTTTGTTCAAAAGTGTACTCGCCCTCCGCTACACCTAACCTAACCTCAGCTCAT 5202
QY 421 TTTAACTGTAATAAAAAACGGCGCTTCAGGTGCAAGTTTATTTATATATATTCCTGCTGGC 480
DB 5203 TTTCAATGTAATAGAGAACAAAGCTTCAGGTGCTACTTTATTCGTTATACATACCAGCTGT 5262
QY 481 GAACATAAAATTTGCTCTTTTGGTGTATCTGGGATGCTACTCTGAAGTTAAGAGTAAAA 540
DB 5263 GAATTAATAAATTTACCTTTTGGGGGGGTCTGGAATGCCGTCTGGAAGCTAAATGTAATA 5322
QY 541 AGACGATAGTCAGACCTATGGAATCTTACATATAAATATACATATTAATTAATTAACCTGAT 600
DB 5323 AGACGATAGTATCAACCTATGGAATCTTACATATAAATATACATATTAATTAATTAACCTGAT 5382
QY 601 AAGGGAATATTCAGATATGTTTACCTCAGTTCAAAAGTACCGCTCCGCTCGATCTTAAC 660
DB 5383 AAGGGAATATTCAGATATGTTTACACACTTCAAAAGTACCGCTCGTCTCGATCTTAAC 5442
QY 661 TTGCGTCCAACTGGTGGGGCACAATATATTGGAAGAAATTTCTGTTGATATGTCCTTTAT 720
DB 5443 TTGCGTCCAACTGGTGGTGTACATATATGCGGAAGAAATCTGTTGATATGTCCTTTAT 5502

```

```

QY 721 GATGGATATAGTACTAACAGCAGCTCTTTGGAGATAGATTTTCAGGATACCAATCTCTAAA 780
Db 5503 GATGGATATAGTACTAACAGCAGCTCTTTGGAGATAGATTTTCAGGATACCAATCTCTAAA 5562
QY 781 TCTGTGGGGAATTTTATCTAGGAAATAAATGATGACACCAAGAAATGCTATATCT 840
Db 5563 TCTGTGGGGAATTTTATCTAGGAAATAAATGATGACCTCCAAAGAACTTGATATACCT 5622
QY 841 TTGTCACTCTCTTGGCGGGTAAAAAGTTTAACTCCCAAAATGGAACGTCAATTAATAAT 900
Db 5623 TTGTCACTCTCTCGGAGGTAAAAATTTAAACACCAAAATGACAGGCAATTAATAT- 5681
QY 901 GTGAGCGAGCTTCTCTGGAACCAAACTGGAATAGAAATTCACCTGTACCATGCCAGAA 960
Db 5682 --TAACTACTCTCTCTCGGAACCAAACTGGAATAGAAATTCACCTGTACCATGCCAGAA 5739
QY 961 ATCAGTGTCCCGGTGTGTGCTGCTGCGCTGGACGTTTCAATTGGATGCAAAAGTGGAAAT 1020
Db 5740 ATCAGTGTCCCGGTGTGTGCTGCTGCGCTGGACGTTTCAATTGGATGCAAAAGTGGAAAT 5799
QY 1021 CCGAGGCTGCAATATATGGGTAAATTAATGTTACTTTCACACCAAGTACTCAAAACA 1080
Db 5800 CCGAGGCTGCAATATATGGGGAATATTAATTAATTTACTTTCACACCAAGTACTCAAAACA 5859
QY 1081 CTCTAG 1086
Db 5860 CTCTAG 5865

RESULT 6
LOCUS      EColi cooC and cooD genes.
DEFINITION E.coli cooC and cooD genes.
ACCESSION  X76908
VERSION    X76908.1 GI:488735
KEYWORDS   cooC gene; cooD gene.
SOURCE     Escherichia coli
ORGANISM   Escherichia coli
REFERENCE  1 (bases 1 to 5336)
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
AUTHORS    Froehlich,B.C., Karakashian,A., Melsen,J.R., Wakefield,J.C. and
            Scott,J.J.
TITLE      CooC and CooD are required for assembly of CSI pili
JOURNAL    Mol. Microbiol. 12 (3), 387-401 (1994)
MEDLINE    94344028
PUBMED     7915003
REFERENCE  2 (bases 1 to 5336)
            Scott,J.J.
TITLE      Direct Submission
JOURNAL    Submitted (24-DEC-1993) J.J. Scott, Emory Uni. School of Medicine,
            Dep. of Microbiol. and Immunol., 1510 Clifton Rd, Atlanta 30322,
            Georgia, USA
COMMENT    Entry cited as #76908 in the journal.
FEATURES   Location/Qualifiers
            source
              1..5336
                /organism="Escherichia coli"
                /mol_type="genomic DNA"
                /strain="LMC10"
                /db_xref="taxon:562"
                /clone="pEU494"
              76..79
                /citation=[1]
              95..2703
                /gene="cooC"
              85..2703
                /gene="cooC"
                /citation=[1]
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CooC"
                /protein_id="CAA54225.1"

```

```

/db_xref="GI:488735"
/db_xref="SPTREMBL:Q47114"
/translation="MIGGSKWIVLVLVIGSSGFASKYNLVIDPESFRDLWGEQD
ELLVRLVQCQSLGVRHRIKSTPTTVAPEPNDLLDKIEINKHEADLRVLMRGSFORNG
NMSCOGYQONNKNYIKNTAVAVDDVENVLNLFIENFLASGENDSYQYPSKNTK
KAFIHSOTINLSDTGNVENLSIVGSLGIDTNSVAILLGAANYRYKSYNNEOSIN
SLYFPHDEKFNYYQI.GRIDSDLSQSGGNGENFDLLPVVDIYGRAGTTQSYIKNTG
KSAVFETIMLTHFSRVEAYRQQLIGVYLDAGISELTERDPDNGYDLKLKIFEQE
QIVRBEI VPFNKGSSIGDTHWDFVQAGDIINDNGRYVEKQKNHKSAINSLGRPLT
RNLVQLGAGVIDNKNYETEGLMNSGLDLSKSTFLFGDTHGNTQNTSYTDGF
SISVHNDKRVDDCGKDNMGWGCYVESASLSIPVKGNSTLAYSNYTSVRYD
AVSVHPVYVYKGTQKQOLTASTVVRMGDYNIMPTIGVNSBQKQWADKGGVLSLTL
TRVDGKSLNAGSYNYSRGNTSNDAFVEGHLVSDTNYVSRELSARVSGNSVYTBGG
TRVGNRRRGDNGTLNKNKRSKSHDTTSLTAGYSSPALTDDGLIYWGSSASGLTNG
SGEIVRYKSNEDSELNLYKSGSYGHSYLSNDSLIFIPVPAKQASLTLEENTKSKN
IDVLAPTNTFFMLPGSVYPIDVSNVSTYVGRGVDKGRPLSGAYIINAQNIVLDE
NGGFGFESSNEKELFLIKDKTIYSCSLDRSEMRNGIAFVGEVACNSTIKELLPEKL
TNSRHDLLAYNQDTE"
83..156
/gene="cooC"
/citation=[1]
157..2700
/gene="cooC"
/product="CooC"
/citation=[1]
2889..2892
/gene="cooC"
/citation=[1]
2700..3791
/gene="cooD"
2700..3791
/gene="cooD"
/citation=[1]
/codon_start=1
/evidence=experimental
/transl_table=11
/product="CooD"
/protein_id="CAA54230.1"
/db_xref="GI:488737"
/db_xref="GOA:Q47115"
/db_xref="SPTREMBL:Q47115"
/translation="MKKIFILSIIFSAVVSAGRYPETTVGNLTKSFQAPRLDRVQS
PIYNTNHWAGYLSHSLYDRIVFLCTSSNPVNGACPTIGTSVGYQYGTITTLQFT
EKRLIKRNINLGNKKPIWENQSCDFSNLWLNKSWSCGAGNAGTLNLIYIPAG
EINKLPFGIWEATLILRLSRYGEVSTHGYNTVINITVDLTDKAGNIQWLPHFNHP
RVDNLRLPIGNKYSGNSLDMCFYDGYTSDSMVVKFQDDPTNSSEVNLKYIGST
EKLFPYAVSLIMGEKIFYPVNGQSTINDSSVLETNWNRVTAVAMPEVNVPLCWPAREL
LLNADVNAPDAGQYSQIYITFTPSVENL"
2700..2753
/gene="cooD"
/citation=[1]
2754..3758
/gene="cooD"
/product="CooD"
/citation=[1]
4476..4536
/notes="bases 1-887 of IS2"
/insertion_seq="IS2"

ORIGIN
Query Match      28.6%; Score 310.8; D3 1; Length 5336;
Best Local Similarity 58.8%; Pred. No. 1.1e-56;
Matches 641; Conservative 0; Mismatches 422; Indels 27; Gaps 5;

QY 7 AAGATTTTATTTTATTTTACATTTGTTTTCTCTTCAGTACTTTTTTACATTTGCTGTATCG 66
Db 2703 AAAAGATATTTATTTTGTCTATCATATTTTCTGCGTGTGCTAGTCCGCGCGATAC 2762
QY 67 GCAGATAAATTCGCGAGATGAAGCATAACTAATATTTTGGCCCGCTGACAGAAC 126
Db 2763 CCGGAAACTACAGTAGGTAATCTGACGAGAGTTTTCAGCCCTCTGCTGGATAGAGC 2822
QY 127 GAATCTTCCCCCAACATAATATATAATAAACATATTACAGCATACAGTGAAGATCAT 186

```

2823	DB	GTACAAATCAACCAATACATAACACATCTTTACGAATCATGTGGCTGGATATAGTTTGGAGTCAT	2882
187	QY	ACTCTCTATCATAGGATGACTTTTTTATGTTTGTCTCTCAAAATACACTTAATGAGCA	246
2883	DB	AGCTTATATGACAGGATGTGTTTTTATGTACATCTCGTGAATCCGGTTAATGTGCT	2942
247	QY	TGTCCAACAGTCAGAAATCCTACAGCTTCATCGGTGAGCGGTGAACAATAATACATTA	306
2943	DB	TGCCCAACCATTTGGAACATCTGGAGTTCAAT-----ACGGTACTCAACACCAATAACCTTTG	2996
307	QY	CAATTTACGGAAAAAGAACTTTAATAAAAAGAGAGCTACAAATTAAAGGCTATAAACAA	366
2997	DB	CAGTTTTACAGAAAAAGAGCTCTGATPAAAAGAAATATTAATCTTCAGGTAAATAGAAA	3056
367	QY	TTATCTGTTCAAAAGTGTAACTCCCATCCCGGCTTAACACTT-----AACTCAGCTCAT	420
3057	DB	CCAAATATGGAGAAATCAGAGTTCCGACTTTAGCAATCTAATGTGTGTGAATTCGAAGTCT	3116
421	QY	TTTAACTGTATAAATAACG--CGGCTCAGGNGCAAGTTTATATTATATATATCTCGTCT	477
3117	DB	TGGAGCTGTGGGCTTACGGAAATGCTAACCGAACTTCTTAATCTGTATATATCCCTGCA	3176
478	QY	GGCGAACTAAAAATTTGCCTTTTTGTGTGATCTGGATGCTACTCTGAAGTTTAAGAGTA	537
3177	DB	GGAGAAATCAACAAATCGCTTTTGGAGGATATGGGAGGCAACTCTGATCTTTACGCTTA	3236
538	QY	AAAG-----ACGATATAGTGAGACCTATGGAACTTACACTATAATATCACTATT	588
3237	DB	TCAAGATPATCGGAAAGTCAGTAGCACCCATACGGCAATATATACCGTAAATATATACGGTT	3296
589	QY	AAATTAACCTGATAAGGGAATATTCAGATATGTTTACTCAGTTTCAAAAGTCAGCTCGC	648
3297	DB	GAITTTAACTGATAAAGTAAATATTCAGTATGGCTTCCAGGGTTTCACACCAACCCGGT	3356
649	QY	GTCAATCTTAACCTTGCTCCAACTGTGTGGGGGCACATATATTGGAAGAAATTCGTGTGAT	708
3357	DB	GTAGACGTGAATCTGCCCTTATCGGTAAATATAAATATATAGTGTGTAGTAAATTCACTCGAC	3416
709	QY	ATGTGCTTTTATGATGGATATAGTACTAACAGCAGCTCTTTGGAGATAAGATTTTCAGAT	768
3417	DB	ATGTGTTTCTATGAAGATATAGTAPCAACACAGTGATAGCATGGTAAATPAAAGTTCCAGAT	3476
769	QY	AACAATCCTAAAICTGTATGGCAATTTTATCTAAGGAAAAATAATGATGACCAACAGAA	828
3477	DB	GAATAATCCTACCAATTCATCTGAATAATACTTTATTAAGATA--GGGGGCACGTGAATA	3533
829	QY	ATTGCAATATACTTGTCACTTCTCTGGCGGTAAAAGTTTAACTCCAAACAAATGGNAAG	888
3534	DB	TTACCATATGCTGTTCACGTCTTATGGAGAAAAATATTTTATCCAGTGAATPGTCAA	3593
889	QY	TCATTAATATTTGTGTGACGAGCTTCTCTGGAAACAAACTGGAATAGAATATACAGCTGTC	948
3594	DB	TCATTTACTATCAATGACAGTAGTGTACTCGAAACAAACTGGAATCAGTAAACCGCAGTT	3653
949	QY	ACCATGCCAGAAATCAGTGTTCGGTGTGTGTGGCTGCAAGTTTGCAATTTGGATGCA	1008
3654	DB	GCTATGCCGGAAGTTAATGTTCACGATATATGTGCGCCACGAAGTTGCTATTAAATGCT	3713
1009	QY	AAAGTGAAAAATCCCGAGGTGGCAATATATGGGTAAATTAATGTACTTTTACACACCA	1068
3714	DB	GATGTAATGCTCCCGATGCGAGACAGTATTCAGGACAGATATATATATATACATTTACACCC	3773
1069	QY	AGTAGTCAAA	1078
3774	DB	AGTGTGAAA	3783

RESULT 7			
AX741413			
LOCUS	AX741413	5336 bp	DNA - linear
DEFINITION	Sequence 3 from Patent WO03022306.		
ACCESSION	AX741413		
			PAT 29-MAY-2003

```

VERSION
KEYWORDS
SOURCE
ORGANISM
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.

REFERENCE
1
AUTHORS
Turner,A.K., Greenwood,J., Stephens,J.C., Beavis,J.C. and
Darsley,M.J.
TITLE
Bacterial vaccine
JOURNAL
Patent: WO 03022306-A 3 20-MAR-2003;
JOURNAL
Scambis Research Limited (GB)
FEATURES
Location/Qualifiers
1..5336
/organism="Escherichia coli"
/mol_type="genomic DNA"
/db_xref="taxon:562"
85..2703
/notes="unamed protein product; cooC"
/codon_start=1
/transl_table=11
/protein_id="CAD90906.1"
/db_xref="GI:30524204"
/db_xref="RENTREMBL:CAD90906"
/translation="XIGGKSSKRVIVISVLIGSSSGFASKYNLVIDPESPRDLWGQDP
ELLEVLQYCGGSHVRIKTNPTTAVFESPDNLDKIEINKGKADRLVMRGFSQNG
NMCCQYCTQCNKVIKTNIVAVDDVENLNLFTIGNEFPLASGENDSDYIQESKNT
KAPFHQTNLSDTQYENLSTIVGTSLGDTNSYAILGWAANYRYKSYNVEQNSIN
SLVFRHDPKNTYQGLDRSDLSQSSGDNFNLDLPVDIYMRAGTIVQSYIKNTG
KSVASPTIMLPHSPSVAYRGQLLGVYLDAGISELDTERLPDGNVDLKLIPQES
QLVREIIVFNGKSSIGDTHMVPQAGDIDNNGRYVEKQKHKASINSGVRLPIT
RLNVLAVGADGKSNYITGILWNSSLGDSLNKSTFLFGDDTHGNTQNSVTLTGDF
SISFPHXRVDDCGDQWNGWSGCYSEYASLSIPVKNWSTLAYSNITSTGVRYD
AYSEVYVYIKRTRKWTASTVVRWGDYNIMPTIYVNSBQKQWADKGGYLSLT
TAVDGKSLNAGSYNYSNGVNTSNDAFVEGHLVSDTNVYSRELSARVSGNRYTGG
VSGINRNRTDGLNLVNKNKSHDTTSLTAGYSSSFALTIDGLIWGGSASGLNKG
SGGIVRKNEDRSNLNKGSSYSHYSIGSNDLSLIPFALMQASLTTEENTKSKN
IDVLATKNTFFMLPGSVYIPDVSANVSFTYVGRVDVGRPLSGAYILNAQNLVLD
NGFASVSENEKEFLDKDTIYSCSLDRSEMRNGIAFYGEVACNSTTIKEILLPEKLV
TNSRIHDLIAYNQDTE"
2700..3791
/notes="unamed protein product; cooD"
/codon_start=1
/transl_table=11
/protein_id="CAD90907.1"
/db_xref="GI:30524205"
/db_xref="RENTREMBL:CAD90907"
/translation="MKKIFILFIISAVVVSAGRYDETVEGNTLKGFQAPRLDRSVOS
PIYNITNHHVAGLSLSHSLYDRIVFCTSSSNPVNGACPTIGTSQVQYGTITTLQPT
EKSLIKRINLAKNPKIWNQSCDFSNLWLNKSWSCGARGNAGTLLNLIYIPAG
EINKLPFGGWEATLILRLSRGYCEVSTHYGNVTNITVBLDKGNIQWLPQFHNPF
RVDLNLRPTGNYAYSGNSLDMCFYDGYSTNSDSMTVKFQDDNPTNSSEYNLYKIGT
KPLPVAVSLMGKRLIPIYPYNGQSFITINDSSVLETNNRVTAVAMPEVNVFLCWPAEL
LMLADPNVLLPAGVQISQIYIIFPSVENL"

ORIGIN
Query Match 28.6%; Score 310.8; DB 6; Length 5336;
Best Local Similarity 58.88; Pred. No. 11e-56;
Matches 41; Conservative 0; Mismatches 422; Indels 27; Gaps 5;

Qy 7 AAGATTTTATTTTACATCTGTTTTCTCTCAGTACTTTTTTACATTGCTGTATCG 66
Db 27C3 AAAAAGATATTATTTTTTTGTCTATCATATTTCTCGGTGCTCAGTCCGGCGGCGATAC 2762

Qy 67 GCAGATAAAATCCCGAGATGAAGACATAACTAAATATTTTGGCCCGCGCTGACGGAAC 126
Db 2763 CCGGAAACTACACTAGTAGTAATCTCACGAAGAGTTTTCGAAGCCCTCGCTCGGATAGAAGC 2822

Qy 127 GAATCTTCCCCCAACATAATATATTAATAACCATATTACACATACAGTGAAGATCAT 186
Db 2823 GTACAATCAACCAATATAAACAATCTTTACGAATCATGTGCTGGATATAGTTTGATGTCAT 2882

Qy 187 ACTCTGTATGATAGGATGACATTTTATTTAIGTTTGTCTCTCAACATACACTTAATGGAGCA 246

```


SGGIVRVKSNEDBSLNNVKGSSYNYSLGNSLSLIPVLPALMOASLTIKENTNKSXN
 IDVLAPKNTFFMPLPGSVPIPDVSANVSFTVVGKGVKGRPLSGAVILNQNVLDE
 NGSFSSSENKEBLFLDKRTIYSCSLRSEMRNGIVFVGEVACNSFKILLPEKLV
 TNSRIHDLAVNQUTE"
 /gene="csbb"
 /db_xref="GI:28932778"
 /translation="MKKIFILFSLIFSVAVSAGRYPEVTVGNLTGKSFQAPQDRSVQS
 PIYNIPTNVAQYSLSHLYRIPIVFLSSPNVNGACPNRGTSQVQGTITTLQFT
 EKRSLIKENILAGNKKRIWNSQSDTSLNVLNKSWSGANGKANGTILNLIYPAG
 EINKLPFGGIWEATILILSLRYGEVSSYGNVTYNIIVDLTDKNGICVWLPGFHNP
 RVDLNLHPIGNKYSGNSLDMCFYDGYSTNSDSNLIKFOEDNPYSSEVLYKIGCT
 EKLPAVSLLMGEKILFHPVNGQSFTINDSSVLEITNWNRTVAVMPVNPVLCWPARL
 LLNADVNAPDRAGQVSGQIYITFTPSVNI."

ORIGIN

Query Match 28.5%; Score 309.6; DB 1; Length 6956;
 Best Local Similarity 58.2%; P: 0; Mismatches 429; Indels 27; Gaps 4;
 Matches 634; Conservative

QY 7 AAGATTTTATTATTATTTTACATGTTTTCTCTTCAGTACTTTTTTACATTTGCTCTATCG 66
 DB 4124 AAAAGATATTATTATTGTTGCTATCATATTTCTGGGTGTCAGTCCGGGGGATAC 4183
 QY 67 GCAGATAAATTCGGGAGATGAAGCATACTAATATTTTGGCCCGGTGACAGGAAC 126
 DB 4184 CCGGAATACAGTAGTAGTAATCTGACGAGAGTTTCAAGCCCTCGTCAGGATAGAAGC 4243
 QY 127 GAATCTTCCCAACACATATATTAATAACCATATTACAGCATACAGTGAAGTCTAT 186
 DB 4244 GTAACATCCGATATATAACATCTTTACGATCATGTGGCTGGATAGTTGAGTCAAT 4303
 QY 187 ACTCTGTATGATAGGATGACATTTTATGTTTCTTCAAAATACACTTAATGAGCA 246
 DB 4304 AGCTTATATGACAGGATTTGTTTTATGTACATCTCTGTGAAATCCGGTTTAAATGTTGT 4363
 QY 247 TGTCCAAACAGTGAATCTTACGAGTTCATCGGTGAGCGGTGAACAAATATACATTA 306
 DB 4364 TCCCAAAACCGTGGACATCTGAGTTCAT-----ACGGTACTACACCAATACCTTG 4417
 QY 307 CAATTTACGGAATAAAGAGTTTATATAAAGAGAGCTACAAATTAAGGCTATATAACAA 366
 DB 4418 CAGTTTACAGAAAAGAGTCTGATATAAAGAAATATTAATCTTGCAGGTATTAAGAA 4477
 QY 367 TTAATGTTTCAAAAGTGTAACTGCCCATCCGGCTTAACACTTAACCTAG-----CT 417
 DB 4478 CCAATATGGGAGAAATCAGAGTTTCGACACTGACAAATCTAATGGTGTGAATTCGAAGTCT 4537
 QY 418 CATTTTAAGTAAATAAACCAGCTTCAGGTGCAAGTTTATTTATTTATTTATTTCTGCT 477
 DB 4538 TGGTCTGTGGGGCTAACCGAAATGCTAACGGAACACTTCTAATCTGTATATCCCTGCA 4597
 QY 478 GCGGAATAAATAATTTGCTTTTGGTGTCTGGAATGCTGAGTGAAGTAAAGATTAAGATTA 537
 DB 4598 GGAGAAATCAACAAATTTGCTTTTGGAGGATATGGAGGCACTCTGATCTTACGCTTA 4657
 QY 538 AAAAG-----ACGATATAGTCAGACCTATGGAACCTTACACTATAAATATCACTATT 588
 DB 4658 TCAAGATATGGCAAGTCACTAGCACCCATTACGGCAATATACCGTAAATATTACGGTT 4717
 QY 589 AAATTAAGTATAGGGAATTTTCAGATATGTTTACCTCAGTTTCAAAAGTACCGTCCG 648
 DB 4718 GATTTAACTGATAAAGGTAATATTACGATATGGCTTCCAGGTTTACAGCAACCCGGCT 4777
 QY 649 GTCGATCTTAAGTGTGCTCCACTGTGGGGGCACATATATTGGAAGAAATTTCTGTGAT 708

DB 4778 GTAGACCTGAATCTGCACCCCTATCGGTAATATAAAATATAGTGGTAGTAAATTCACCTCGAC 4837
 QY 709 ATGTGCTTTTATGATGATATAGTACTAAACAGCAGCTCTTTGGAGATAGATATTTCCAGGAT 768
 DB 4838 ATGTGTTTCTATGATGATATAGTACTAAACAGTGTATAGCATGGTAAATAAGTTCAGGAT 4897
 QY 769 AACAAATCCTAAATCTGATGGCAAAATTTTATCTAAAGAAAATAAATGATGACACCAAGAA 828
 DB 4898 GATAATCCTACTATCTATCTGAATATATATTTTATAAGATA--GGGGGCACTGAAAA 4954
 QY 829 ATTGCATATCTTTTGTCACTTTCTCTTCGGGGTAAAAAGTTTAACTCCAAACAAATGCAAG 888
 DB 4955 TTACCATATGCTGTTTCACTCTTATGGCAGAAAAATAATTTTCACTCCAGTGAATGCTCAA 5014
 QY 889 TCATTAATATTTGTCGACGAGCTTCTCTGAAAACAAACTGGAATAGATTAACACTCTC 948
 DB 5015 TCATTTACTATCAATGACAGTAGTGTACTCGAAAACAAACTGGAATCGAGTAAACCCAGT 5074
 QY 949 ACCATGCCAGAAATCAGTGTGTTCCGGTGTGTTGCTGCTGACAGCTTGCATTTGGATGCA 1008
 DB 5075 GCTATGCCCGAAGTTAATGTTCCAGTATATGCTGCCGACAGATGCTGTTATTAATGCT 5134
 QY 1009 AAAGTGGAAATCCCGAGGCTGCAATATATATGTTGTTTAACTTTTACTTTTCAACCA 1068
 DB 5135 GATGTAATGCTCCGATGTCAGGACAGTATTTCAGGACAGATATATAACATTTACACC 5194
 QY 1069 AGTAGTCAAA 1078
 DB 5195 AGTGTGCAAA 5204

RESULT 9
 ECOTABCD
 LOCUS E.coli cotA, cotB, cotC and cotD genes.
 DEFINITION
 ACCESSION Z47800
 VERSION Z47800.1 GI:897725
 KEYWORDS cotA gene; cotB gene; cotC gene; cotD gene.
 SOURCE Escherichia coli
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.
 REFERENCE 1 (bases 1 to 5798)
 AUTHORS Froehlich, B.J., Karakashian, A., Sakellaris, H. and Scott, J.R.
 TITLE Genes for CS2 plli of enterotoxigenic Escherichia coli and their
 interchangeability with those for CSI plli
 JOURNAL Infect. Immun. 63 (12), 4849-4856 (1995)
 MEDLINE 96071908
 PUBMED 7591145
 REFERENCE 2 (bases 1 to 5798)
 AUTHORS Froehlich, B.J.
 TITLE Direct Submission
 JOURNAL Submitted (18-JAN-1995) Froehlich B. J., Emory University,
 Department of Microbiology and Immunology, 1510 Clifton Rd, Atlanta,
 GA, USA, 30322

FEATURES
 source
 1..5798
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /strain="C9if-6"
 /db_xref="taxon:562"
 /clone="pEJ5C06"
 1..369
 /note="99% homologous to last 369 bases of IS3"
 /citation={1}
 /label=IS3
 369..401
 /note="100% homologous to bases 73-41 of IS1"
 /citation={1}
 /label=IS1
 422..427
 /label=IS1
 446..451
 /citation={1}

misc_feature
 1..369
 /note="99% homologous to last 369 bases of IS3"
 /citation={1}
 /label=IS3
 369..401
 /note="100% homologous to bases 73-41 of IS1"
 /citation={1}
 /label=IS1
 422..427
 /label=IS1
 446..451
 /citation={1}

-35_signal
 -10_signal

```

RBS      495..488
/citation=[1]
gene      499..1215
/gene="cotB"
CDS      499..1215
/gene="cotB"
/citation=[1]
/codon_start=1
/transl_table=11
/product="CotB"
/protein_id="CAA87760.1"
/db_xref="GI:897726"
/db_xref="SPTREMBL:Q47116"
/translation="MKILLFVLFNFVPAASANFVPIPSKIDIQSGSETIKVFKSK
DVQYIKYIKRVINPTKEEQEVDIKMDGGLIVTPAKVLPAGASKRIURIEINKKE
QEEVRYVESVPGQODDIEEKGRVNTDLVNIIYAALIRITSPENPQKIDVSIIES
NNWIKNTGNIKILGKIDVFLCDTTSINDKCAKFSYRNLYPDMGVDTKLKGGKFSYAV
IDTKDRNENSGELINIKLP"
sig_peptide      499..552
/gene="cotB"
mat_peptide      553..1212
/citation=[1]
/gene="cotB"
/product="CotB"
/citation=[1]
1241..1245
/citation=[1]
1255..1767
/gene="cotA"
1255..1767
/gene="cotA"
/citation=[1]
/codon_start=1
/evidence=experimental
/transl_table=11
/product="CotA"
/protein_id="CAA87761.1"
/db_xref="GI:897727"
/db_xref="SPTREMBL:Q47117"
/translation="MKLNKIIIGALVLSSTFVSGASAEKNIIVTASVDPTIDLQSQD
GTALPSAVNIAYLPQEKRPESARINQVHTNNKTIQIKLINDNVVMTNLSDPKRTI
PLEVSPAGTKLSTAITSITADQLNFGAAGVETVSATKELVINAGSTQQTNIVAGNYQG
LVSIVLTQSP"
sig_peptide      1255..1323
/gene="cotA"
/citation=[1]
/evidence=experimental
1324..1764
/gene="cotA"
/product="CotA"
/function="major pilin subunit for CS2 pili"
/citation=[1]
/evidence=experimental
1822..1824
/citation=[1]
1836..4436
/gene="cotC"
1836..4436
/gene="cotC"
/citation=[1]
/codon_start=1
/transl_table=11
/product="CotC"
/protein_id="CAA87762.1"
/db_xref="GI:897728"
/db_xref="SPTREMBL:Q47119"
/translation="MRAFNKITYVILFIPGLCFGTNGLBSKKNIPSEFIDLMWEODEL
LEVNLRSLGVRHVLITPTTVFSSVEEILEKINVKQEKEDLSLILQSYSENGM
SCNGFEKRYSCNIRITDVNVIIDENNELNFIIGASFLSVQAOQNIITYQKNINEK
AFIESQTINPESSEGYKSLKGVGAQGLTENSYLVFGMDAIYNSRKYTYIKQNSINN
IYRVDFDKLYYYQLGMDRSLSSASGNFNFMPLPDPIDGFIQIGTQQSYIKNIEK
SISSPVTVMLTFRSRVAPRNBELLGWYLNLSGINDLD"SRLPDGSYDLTLKIFEQD
LVREKVPFNKGGASPDQXQDVPFAQAGNIVNNNDSVIERKQTKNKTGINAGITPVTR

```

```

NLSTLQGGAIIDNDKYEAGVNRSGFLDGLVLSGNFSLYGDGARGNYQNIISYTDGPN
LSFYENDKSVNDCSHNTSACHSGCYESYPSLSVPSWCTTTLGVNHTNNEAVHKVDY
TPEIFFSKYIGVSRWOLITSSSIKWMMDIYVIFITGYRSDQSRWSQGGYFSLSFT
RVKNSAINAGYSYNYVKHNATHEAFUDGRITNTNTFGYSELGRINTKNNTEAGVT
GRVKNRFDGLNGSLNVNKSKTSGKWTSHMSANSYNSFAITGDSVTWGGDAGLTKLSG
GVNVRSDDKKELIKIGSSSYNYILGSDNDRSFTIPVSALMPSNLITIEIQSDKNVIT
VQALSKNDFPILPGNVFPIDVTANVTYSYIGRALDDKGNPLSNAILDVGVRIDEDG
GFSPTSQAQKSLFLFKODIYSCDVKKYDLRSGLVLTGDLICEHSGIERLQKDLVNN
PRVQLLJYK"
sig_peptide      1836..1913
/gene="cotC"
/citation=[1]
1914..4433
/gene="cotC"
/product="CotC"
/citation=[1]
4441..4445
/citation=[1]
4451..5545
/gene="cotD"
4451..5545
/gene="cotD"
4451..5545
/gene="cotD"
/note="ttg start codon"
/citation=[1]
/codon_start=1
/transl_table=11
/product="CotD"
/protein_id="CAA87763.1"
/db_xref="GI:897729"
/db_xref="SPTREMBL:Q47119"
/translation="MKKVIIVLSMFLCQVYQSWHTNVEAGSINKTESIGIDRSAA
ASYPAEYIFHEHVAGYKDHSLFDRMTFLCMSSTDASKACPTGNSKSSQGETNKL
IPEKSLARKTILNKGKRYELYESDRCLIHVYDKNLNSHTVKYGSFTRGVDFLYI
PQGEIDGLLTGGIWEATLELRVHRHYDTHNGYKKNIVFDLTDKNGNIQWTFKPSDP
RIDNLRPEGNGKSGSNVEMCLYDGYSTHSQS:EMRPQDDSCQTGNNEINLIKXGEP
LKLLPYKLSLIGREFYPNNGEAPTINDTSSLFINWNRIRKSVSLPQISIPVLCWFPAN
LTEMSELNPNPAGYSGILNVVTFPSSSL"
sig_peptide      4451..4504
/gene="cotD"
/citation=[1]
4505..5542
/gene="cotD"
/product="CotD"
/citation=[1]

```

ORIGIN

```

Query Match      27.5%; Score 298.8; DB 1; Length 5798;
Best Local Similarity 58.5%; Pred. No. 4.1e-54;
Matches 566; Conservative 0; Mismatches 387; Indels 15; Gaps 2;

QY      127 GAATCTTCCCCCAACATAATATATAATACCAATACCATATACAGATACAGTGAAGTCAT 186
DB      4583 GATCGTATCCCTGCTCAATTATATATTCATGACATGTTGCGGTGGTACAAATAGATCAC 4542

QY      187 ACTCTGTATGATAGGATGACTTTTTTATGTTTGTCTCTCACAAATACACTTAATGGAGCA 246
DB      4643 TCCTCTTTTCAGAGGATGACGTTTATGATGTCATCAACAGATGCATCTAAAGTGCA 4702

QY      247 TGTCCAAACAGTGAGAAATCCTAGCAGTTCATCGGTGAGCGGTGAACAATATATAATTA 306
DB      4703 TGTCCGACAGGAGAAAACCTCCA-----AATCCTCTCAAGGGGAGACTAATATAAGCTA 4756

QY      307 CAATTTACGGAAAAAGAAAGTTTAATAAAAGAGAGCTACAAATTAAGGCTATAAAACAA 366
DB      4757 ATATTTACTGAAAAAGAAAAGTGTGCCCAAGAAAACATTAACATTAAGAGATATAAGAGA 4816

QY      367 TTATCTTCAAAGTGTTAACTGC-----CCATCCGGCCTAACACTTAACCTCAGCT 417
DB      4817 TTTTATATGATCAGATAGATGCATTCATTATGTGATAAAAATGAATCTCAATTCCTCAT 4876

QY      418 CATTTTAATCTGAATAAAAACGGGGCTTCAGGTGCAAGTTTATATTTATATATTCCTGCT 477
DB      4877 ACTGTTAAATGCTAGGTTCAATTCACAGAGGAGTAGATTTTCATTTATATATATCCACAA 4936

```


QY	478	GGCGAACTAAAAAATTTGCGTTTGGTGGTATCTGGGATCGTACTCTGAAGTTAAGAGHA	537
Db	4937	GGTGAATTTGATGGGCTTCTAACTGGAGGTATATGGGAGCAACACTAGAGTTACGAGTC	4996
QY	538	AAAAGACGATATAGTGAGACCTATGGAACCTTACACTATAATATCACTATTAAATTAAC	597
Db	4997	AAAAGCAATTCAGACTATAATCTAAGGACHTTCAAGGTTAATATCACTAGTTGATTTGACA	5056
QY	598	GATAAGGAAATATTCAGATATATGTTACCTCAGTTCAAAAGTGACGCTCGCGTCATCTT	657
Db	5057	GACAAAGGAAATATTCAGGCTCGGACACCAAGGTTTCATAGCGATCTAGAAATGATCTG	5116
QY	658	AACCTTCGTCACACTGCTGGGGGACATATATTTGGAAGAAATCTGTGTATGTGCTTT	717
Db	5117	AAVTTACGTCCTCAAGGTAATGTTAAATATTTCTGGTAGTAACGCTGTGATGTGCTC	5176
QY	718	TATGATGGATATAGTACTAACACAGCCTCTTTGGAGATAGATTTCAAGGATAACAATCCT	777
Db	5177	TATGATGGCTATAGTACACATAGTCAAAGTATAGAAATGAGGTTTCAGGATGACTCACA	5236
QY	778	AAATCTGATGGGAAATTTTATCTAAGGAAATATAATGATGACACCAAGAAATTTGCATAT	837
Db	5237	ACAGGAAATATCAATATATCTTATAAAACTGGAGAGCCATTAAAAAATTTGCCATAT	5296
QY	838	ACTTTGCTACCTCTCTTGGCGGGTAAAGTTTAACTCCACAAATGGAACGTCATTAAT	897
Db	5297	AAACTTTCTCTCTTTTAGGAGACGAGAGTTTATCCAAATATGAGAGGCTTTTACT	5356
QY	898	ATTGCTGACGACGCTCTCTCGAAACAACTTGAATGATAGATTTACAGCTGTCAACATGCCA	957
Db	5357	ATTAAATGATCTCTGCTCAATTTGTTATTAAGTAACTGAACTGATTAAGTCTGTATCTACCA	5416
QY	958	GAATCAGTGTTCCTGGGTTTCTGTTGCTCGGACGTTTGCATTTGATGCAATGGAAGTGAA	1017
Db	5417	CAGATTAGTATTCAGTACTATCTGCTGCCAGCAAACTTGACATTTATGTCAGAGCTAAAT	5476
QY	1018	AATCCCGAGGCTGCAATATATGTTGTTAAATTAAGTTACTTTTCACACCAAGTATCAA	1077
Db	5477	AATCCGAGCGGTGAGTATTCAGGAATACTTAAGCAATTTACTCTCTAGTAGTTCA	5536
QY	1078	ACACTCTA 1085	
Db	5537	AGTCTGTA 5544	
RESULT 10			
AX741415			
LOCUS	AX741415	5798 bp	DNA linear PAT 29-MAY-2003
DEFINITION	Sequence 5 from Patent WO03022306.		
ACCESSION	AX741415		
VERSION	AX741415.1	GI:30524208	
KEYWORDS			
SOURCE	Escherichia coli		
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.		
REFERENCE	1		
AUTHORS	Turner, A.K., Greenwood, J., Stephens, J.C., Beavis, J.C. and Darsley, M.J.		
TITLE	Bacterial vaccine		
JOURNAL	Patent: WO 03022306-A 5 20-MAR-2003;		
FEATURES	Acambis Research limited (GB)		
source	Location/Qualifiers		
	1..5798		
	/organism="Escherichia coli"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:562"		
	499..1215		
gene	/gene="cotB"		
CDS	499..1215		
	/gene="cotB"		
	/codon_start=1		

/transl_table=11			
/protein_id="CAD90909.1"			
/db_xref="GI:30524209"			
/db_xref="REMRMBL:CAD90909"			
/translation="MKILLFVILFNVFAASANFWYPIKSDIQSGSETIKVFSKSK			
DVQYIKYIKRVINPQTEQVDIKNWGGGLIVTPAKVLPAGASGSRIRLTINKE			
QREVYVYFESVKGQDDDI EBNKGRVNTDLSNIIYAALIRTPSPNPKRLDVIES			
NRVWIKNTGNIRLGIKIDVFLCDTTSINDKCAKFSYNNL:YFDMSVDTKLGKGFSAV			
IDTKDNRNNSGELINIKLP"			
gene	1255..1767		
CDS	1255..1767		
	/gene="cotA"		
	/gene="cotA"		
	/codon_start=1		
	/transl_table=11		
	/protein_id="CAD90910.1"		
	/db_xref="GI:30524210"		
	/db_xref="REMRMBL:CAD90910"		
	/translation="VKLNKILGALVLSSTFVSMGASAAEKNIITVTASVDPTIDMOSD		
	GIALPSAVNLALPGEKRFESARINQVHTNNTKGLQIKLINDNVNINLSDPSKII		
	PLEVSPAGTKLSATATSIADQLNFGAAGVETVTSKELVINAGTQQQINIVAGNYQG		
	LNSVILVTPP"		
gene	1836..4436		
CDS	1836..4436		
	/gene="cotC"		
	/gene="cotC"		
	/codon_start=1		
	/transl_table=11		
	/protein_id="CAD90911.1"		
	/db_xref="GI:30524211"		
	/db_xref="REMRMBL:CAD90911"		
	/translation="MRAFNKITVFLIFPGLCGTNGLSKKNIPEEFIDLWQBDEL		
	LEVNLXGSLGHRVLTPTTVKFSVEEILEKINVKQEKEDLSLLQSYSRNGM		
	SCNGDFEYSCNRYPTDVTNVIVDEENNELNLFICAGFLSVOAONIYYQKNISEK		
	ATHTIDNFSESGYKSLKGVGAQGLTENSILVFGMDAIYNSRKTYTKQSIIN		
	IYRYPFDKYYQLGRMDRDLSSASGNFNMPLPLPDIGFOCTQOSYIKNTEK		
	SISSPTVMTLTPESVEAPNEELLGVWLNGINDLDSRLPDGSYDILTKI9EQDI		
	LYREEKVPKKGASFGDMQWDVFAQAGNIYNNDSYIEKQTKKZKGINAGITPVR		
	NUSFIQGGAILINDKTYEAGVNRWSGFLDVLGNSFSLYDGGARGNYQISITDGFN		
	LGFYRNDKDSCHNSYAGWGCYESYFSLSVPVSGWTTTLGYNHTNNEAVHKYDY		
	TPEYFPKRYKGVSKRWQITSSSYKMDYHVIPTTIGVYSDQSRSEQGYFSLST		
	RVKENSAINAGYSYVVKHNAIHEAFDGRITNTTGYSELGSRNTNKNNTAGVT		
	GAKVNEFDLNGSLANNVSKTSKMTSHSMANYSFAITGDSVYWGDAAGLTKLGG		
	GVNVRSDDKSRELIKISGSSYGNVILGSDRSFIPVSALMPSNLIIEIQSDKNIT		
	VQALSKNDFILPQNVPIIDVTANVTYSYIGALDDKGNPLSNAILLDVHGVRLDSGG		
	GFSFETSAQKXSLFLKXDXDYSCDVKKYDLRSGLVTGDLTCEHSGIERLGLDLYNN		
	PRVKQLLAYK"		
gene	4451..5545		
CDS	4451..5545		
	/gene="cotD"		
	/gene="cotD"		
	/codon_start=1		
	/transl_table=11		
	/protein_id="CAD90912.1"		
	/db_xref="GI:30524212"		
	/db_xref="REMRMBL:CAD90912"		
	/translation="MKRVIFVLSMFLCSOVYQSWTNEAGSINTKESIGPIDRGA		
	ASYPAHYIPEHVAGNYKDHSLFDRMTFLCNGSTDASKACPTGENSKSQSGETNKL		
	IFTEKKSLARKTNILNKGYRFLYEDRCIHYVDKMNLSHNGHVKCVGSFTRGVDTLYI		
	POGEIDGLTGGIWEATLELRVKRHYDYNHGYKVNITVDLTDKGNIQWTFPHSDP		
	RIDNLNRPNGKNGYSSNVLEMLYDGYSTHSQSIEMRFODDSQTCNNYNNLIKTEGP		
	JXKLPYKLSLLGGREFYVNGEAFINDTSSLFINWNRKXSVSLPQISIPVLCPAN		
	JTFMSELNPNPEAGYSGILNVTFTFSSSL"		

ORIGIN

Query Match	27.5%	Score 298.8	DB 6	Length 5798
Best Local Similarity	58.5%	Pred. No. 4.1e-54		
Matches 566	Conservative 0	Mismatches 387	Indels 15	Gaps 2
QY	127	GAATCTTCCCCCAACATAATATTAATAAATACCATATTACACATACAGTGAAGTCAT	186	
Db	4583	GCATCTGATCCCTCTCATTAATATATTTTCATGAACATGTTGCTGGTTACAATAAAGATCAC	4542	

187 ACTCTGTATGATGAGTACCTTTTATGTTTGTCTTCTCAAAATACACTTAATCGAGCA 246
Db TCTCTTTTTCAGAGTACGCTTTTATGTTTGTCTTCTCAAAATACACTTAATCGAGCA 4702
247 TGTCCAAACGATGAGAACTCTAGCAGTTCATCGGTCAGCGGTGAGACAAATATACATTA 306
Db TGTCCGACAGGAGAACTCTCA-----AATCCTCTCAAGGGGAGACTATATTAAGCTA 4756
307 CAATTTCGGAAGAAAGAGTGTATATAAAGAGAGCTACAAATTAAGAGCTATATAACAA 366
Db AATTTTACTGAAAGAAAGAGTCTGCCAGAAACATTAATTAAGATATAAGAGA 4816
367 TTAATGTTCAAAAGTGTAACTGC-----CCATCCGGCTTACACTTAACCTCAGCT 417
Db TTTTATATCAATCAGATGAGTCAATCATATGTCGATAAATAATGAATCAATCTCAT 4876
418 CATTTTAACTGTAAATAAAGCGGCTCAGGTGCAAGTTTATATATATATCTCTCTGCT 477
Db ACTGTTAAATGTAGGTTCAATCAAGAGGAGTATGTTTACCTTTATATATCCACAA 4936
478 GCGGAACCTAAATAATTTGCTTTTGGTGTATCTGGGATGCTACTCTGAAGTAAAGATA 537
Db GGTGAATCAATGCGGCTTCTAAGTGGAGTATATGGAGGCAACTAGAGTTACGAGTC 4996
538 AAAAGACGATATAGTGAGACCTATGAACTTACACTATATAATCTACTATTAATTAAC 597
Db AAAAGGCATACCGACTATAATCATGCTACTTACAAAGTTAATATCATCAGTTGATTTGCA 5056
598 GATAAGGGAATATTCAGATAGTTCCTCAGCTTCAAAAGTCAGCGCTCGCTGCAATCT 657
Db GACAAAGGAATATTCAGGCTCTCGAACCCAAAGTTTCAAGCCTCTAGAAATGATCTG 5116
658 AACTTCGCTCCAACTGCTGGGGGACATATATTTGGAAGAAATTTCTGTTGATATGCTTT 717
Db AATTTAGCTCTCAGATATGTTAAATATTTCTGTTAGTAACTGCTGATGCTCTC 5176
718 TATGATGATATAGTAACTGAGAGCTCTTGGAGATAGATTTTCAGGATACAAATCTCT 777
Db TATGATGGCTATAGTACACATAGTCAAAGTATAGAAATGAGTTTCAGGATGACTCACAA 5236
778 AATCTGATGGAAATTTTATCAAGAAATAATATGATGACACCAAGAAATGATAT 837
Db ACAGGAATATAGATATATCTTTAAACCTGGAGAGCCATTAATAAATTTGCCATAT 5296
838 ACTTTGCTACTTCTCTGCGGGTAAAGTTTAACTCCACAAATGGAACGCTCAATTAAT 897
Db AACTTTCTCTCTTTTAGGAGGACGAGAGTTTATCCAAATAATGAGAGGCTTTTACT 5356
898 ATTGCTGAGGAGCTTCTCTGGAACAAACTGGAATAGATTTACAGCTGTCACCATGCCA 957
Db ATTAATGATATCTCTGCTCAATTTTATAACTGGAATCGTATTAAGTCTGTATCTTACCA 5416
958 GAAATCAGTGTTCGGGTGTTCTGTTGCTGCAAGTTTGCATTTGGATGCAAAAGTGGAA 1017
Db CAGATTTAGTATTCAGTACTATGTTGGCCAGCAAACTTGACATTTTCTCAGAGCTTAAT 5476
1018 AATCCGAGGCTGGACATATATGGTGAATATTAATGTTACTTTTACACCAAGTAGTCAA 1077
Db AATCCAGAGCGGTGAGTATTCAGGATATCTTAACGTAACATTTACTCTCTCTAGTATCA 5536
1078 ACATCTTA 1085
Db AGTCTGTA 5544

RESULT 11
AX083744/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
AX083744
Sequence 22 from Patent WO011061.
AX083744
AX083744.1 GI:13185472
synthetic construct

ORGANISM synthetic construct
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1.1141
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
1.1141
/note="consensus sequence of A.t., L.a., and B.n. PAE1
promoters"
ORIGIN
Query Match 5.8%; Score 63.4; DB 6; Length 1141;
Best Local Similarity 10.6%; Pred. No. 0.0019;
Matches 58; Conservative 247; Mismatches 239; Indels 3; Gaps 1;
QY 165 TCACATACACTTAATGAGCATGTCACACAGTGAATCTCAGTTCATCGGCTCAG 284
Db 633 TRMTMRMTMGDGMTVRKKVKWRTTCTTYVDVWADSWVWYANWRCRDVYTRNNYIC 574
QY 225 TCACATACACTTAATGAGCATGTCACACAGTGAATCTCAGTTCATCGGCTCAG 284
Db 573 KSYAISYTWNNWYRYSARNWS---SMARWTRNNWWSGVSVMRWAGTMMWRHN 517
QY 285 CGGTGAAACAAATAACATTTACGGAAGAAAGTATTAATAAAGAGAGCT 344
Db 516 NNNTDTRYWKKWABBTCTTYVDSVCNAKSWRGNNWRAMKMWAAANNNDAGAMDHTY 457
QY 345 ACAATTAAGGCTATAAACAATTTATGTTCAAAGTGTAACTGCCATCCGCTAAC 404
Db 456 MGNNTWWRAMKMMNMAWCCRRAYCCNNNNNNACVWHKHMMRWTKYKWKAAACNNNEK 397
QY 405 ACTTACTCAGCTCATTTTAACTGTAATAAAGCGGCTTCAGGTGCAAGTTTATATT 464
Db 396 AMFMRVAMWYGRDITNTDMMWTSDBWHWTVDYTWKRAVNNNNNNNNWRRBCKTTSM 337
QY 465 AATATATCTGCTGGCAACTAAATAATTTGCTTTTGGTGGTATCTGGATGCTACTCT 524
Db 336 WMDXNTHCTYGNNTWGSAYBYAAMSMAAGASNBVTYKWCWMTYKMTMTNNNNNKA 277
QY 525 GAAGTTAAGAGTAAAGACGATATAGTGAAGCTATGAGACCTATGACATATAATATCA 584
Db 276 WYRTTAVMNNRYTDTAVTBEKRYKYCYAYBYBYBYBYBYBYBYBYBYBYBYBYBY 217
QY 585 TATTAAATTAAGTAAAGGAAATATTCAATATGTTACCTCAGTTCAAAAGTGAAGC 644
Db 216 WWKCRKXVSVHYHAMRYBKWABAVGNNNNKDRMAHHWCAATNNNNMMWVYVMMHM 157
QY 645 TCGCTGCTGATTTAATCTGCTCAACTGGTGGGGGACATATATTTGGAAGAAATCTGT 704
Db 156 HKKGAAATWTKTABRDDHBAHVKTWYWDYWCAMCWMNAKAKVKTAMKEMWYTDY 97
QY 705 TGATATG 711
Db 96 VSANNTG 90

RESULT 12
CEY50E8A
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
CEY50E8A
Caenorhabditis elegans YAC Y50E8A, complete sequence.
AL117200
AL117200.2 GI:6580305
HTG
Caenorhabditis elegans
Caenorhabditis elegans
Bukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.

```

1 none.
AUTHORS Genome sequence of the nematode C. elegans: a platform for
TITLE investigating biology. The C. elegans Sequencing Consortium
JOURNAL Science 282 (5396), 2012-2018 (1998)
MEDLINE 99069613
REMARK The C.elegans Sequencing Consortium.
AUTHORS 2 (bases 1 to 61864)
TITLE Direct Submission
JOURNAL Submitted (06-SEP-1999) Nematode Sequencing Project, Sanger
Institute, Hinxton, Cambridge CB10 1SA, England, and Department of
Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
jesse.sanger@ac.uk or rwem@nematoe.wustl.edu
On Dec 15, 1999 this sequence version replaced gi:5832869.
Coding sequences below are predicted from computer analysis, using
predictions from Genefinder (P. Green, U. Washington), and other
available information.
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
[991122 dnl] : Frameshift detected in sequence. Single nucleotide
removed from the project.
For a graphical representation of this sequence and its analysis
see: - http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?
name=Y50E8A
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we arrange for a small
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
IMPORTANT: This sequence is not the entire insert of clone Y50E8A.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone C48G7 is at 61761 in this sequence. The
true right end of clone R08A2 is at 106 in this sequence. The start
of this sequence (1..106) overlaps with the end of sequence Z98853.
The end of this sequence (61761..61864) overlaps with the start of
sequence Z78061.
FEATURES             Location/Qualifiers
     source            1..61864
                     /organism="Caenorhabditis elegans"
                     /mol_type="Genomic DNA"
                     /strain="Bristol K2"
                     /db_xref="taxon:6239"
                     /chromosome="V"
                     /clone="Y50E8A"
     join(1923..2105,2179..2256)
     /gene="Y50E8A.1"
     join(1923..2105,2179..2256)
     /gene="Y50E8A.1"
     /standard_name="Y50E8A.1"
     /codon_start=1
     /product="Hypothetical protein Y50E8A.1"
     /protein_id="CAB55044.1"
     /db_xref="GI:5832870"
     /translation="WNFIHYILLMSLSISIVTARPPGNDQEKLLTGGEPELGDYV
DYPDFFPVLNKSSTKLNENVQYEKKYLSVAGFPAPKMKML"
     complement(3457..3756)
     /gene="Y50E8A.2"
     complement(3457..3756)
     /gene="Y50E8A.2"
     /standard_name="Y50E8A.2"
     /codon_start=1
     /product="Hypothetical protein Y50E8A.2"
     /protein_id="CAB55045.1"
     /db_xref="GI:5832871"
     /translation="MRLGLVDAPCGVLDPKETVLLDVSRDAFAFGQESTNDRILVWEV
TNKPBAAQFRRGCGQGMTRKNSQLSTHLNKSVPFLVFFFLSNKNSFT"
     join(4673..4783,5619..5699,5756..5833,5909..6094)
     /gene="oig-3"

```

```

CDS

```

```

join(4673..4783,5619..5699,5756..5833,5909..6094)
/gene="oig-3"
/standard_name="Y50E8A.3"
/codon_start=1
/product="C. elegans OIG-3 protein (corresponding sequence
Y50E8A.3)"
/protein_id="CAB55046.1"
/db_xref="GI:5832872"
/translation="MSYSPFPQPQYQSSTLVDCVGVRSQQVSPBLTKOVNHLAY
IPKKNYANNPNWAPFLKVGDLRVHIGKAVDFCVLLGSPRVCWLFNEDKLR
SDVILIEDVADLCRLTIPVQVMEHFGCTFVLCENEVGRATAQAELLPLVE"
complement(6492..9248)
/gene="unc-61"
complement(join(6492..6734,6791..7280,7600..7740,
8092..8173,8224..8363,8412..8506,8622..8789,8840..8893,
9069..9248))
/gene="unc-61"
/standard_name="Y50E8A.4a"
/notes="cDNA EST YK73392.5 comes from this gene
cDNA EST YK73392.3 comes from this gene"
/codon_start=1
/product="C. elegans UNC-61 protein (corresponding
sequence Y50E8A.4a)"
/protein_id="CAD54172.1"
/db_xref="GI:24817603"
/translation="MSFTILITASALLFIFLVLITTAFLIVRSKQSSINLQTVVY
THENPYVTWTSOSSKMDIPVILPRFNMSDIEHKHLPHPQPPVPHHQNQPTNNT
TTISSATSSINTTTSKPTTAAPTAPSPKISLSDHTGLMQLNGHVGDSLPHQLVK
KAVAGAGFNLNMGVETGTGKTLIESLFNKLDEPCNHELKTVELRQCTKDVAEGG
LRVRLRVETAGVFGDQDKTSKAKVIVDLYESQFETYLQELKPRMQLQYFNDLSIHA
CLYFSPGHGLKALDLVTLRLAKRVNVIPIVIAKSDTTCDELLRFAKILSELKSO
KIDITPFDDETSTTKENKNSVPPFAVGSIDPVKENGQMVZARQYPMWIVVEN
ESHQDFVKLRALLTNVDEMKRTHESLYENRRDRKQMKIGDGTGPKLIEKLAQ
KHREQDFSRRLTREFQKLDVTGDMRKVEEGLAREYEHENYREASKLDM
EIRQTEERMKLMIKVSKLARK"
complement(join(6492..6734,6791..7280,7600..7740,
8092..8173,8224..8363,8412..8506,8622..8795,8840..8860))
/gene="unc-61"
/standard_name="Y50E8A.4b"
/notes="contains similarity to Pfam domain: PF00735 (Cell
division protein), Score=415.1, E-value=2.1e-121, N=1
cDNA EST YK109f1.5 comes from this gene
cDNA EST YK22f8.5 comes from this gene
cDNA EST YK411f2.3 comes from this gene
cDNA EST YK524f2.3 comes from this gene
cDNA EST YK411f2.5 comes from this gene
cDNA EST YK510f2.3 comes from this gene
cDNA EST YK510f2.5 comes from this gene"
/codon_start=1
/product="C. elegans UNC-61 protein (corresponding
sequence Y50E8A.4b)"
/protein_id="CAB60583.1"
/db_xref="GI:6425352"
/translation="MSDIEHKLQHPHPQPPVPHHQNQPTNNTTSSATSSIN
TTTTSKPTIAAPTAPSPKISLSDHTGLMQLNGHVGDSLPHQLKAVRAGVQFNL
MCVGTGKTKTLIESLFNKLDEPCNHELKTVELRQCTKDVAEGGIRVRLRVETA
GFGDQDKDKAKVIVDLYESQFETYLQELKPRMQLQYFNDLSIHA
LKALDLVTLRLAKRVNVIPIVIAKSDTTCDELLRFAKILSELKSKIDITPFD
ETVSTTKENKNSVPPFAVGSIDPVKENGQMVZARQYPMWIVVENESHQDFVKLR
ALLRNVDEMKRTHESLYENRRDRKQMKIGDGTGPKLIEKLAQKHREQDFSR
RELTLUREFQKLDVTGDMRKVEEGLAREYEHENYREASKLDMELRQLTRESRK
LMTKYSKLRK"
complement(join(9824..9902,9968..10102,10156..10258,
10304..10469,10683..10728,10840..10979,11047..11151))
/gene="Y50E8A.5"
complement(join(9824..9902,9968..10102,10156..10258,
10304..10469,10683..10728,10840..10979,11047..11151))
/gene="Y50E8A.5"
/standard_name="Y50E8A.5"
/codon_start=1
/product="Hypothetical p-rotein Y50E8A.5"
/protein_id="CAB55048.1"
/db_xref="GI:5832874"

```

```

CDS

```

```

CDS

```

```
/translation="MFLRSIS:LLVLYTYRLTADGYMCKKASDADVWSINKAVGPT
MHVQYSGISVDSINILLISLEIITPYAHSRHAENGPVIVTRCQGETQNMWVKY
GLVNVQGSIVSDDKLVENKIKKQDSCAGYCMINTTRIDEAAGKTIATVRDQV
RTRNLDLLESAPCTVEKLVNRNVNITTCYRGHYCNGLGFAEATRHSKALELP
VNSIDPKSSGPPCTCRKKNREYEMGRKOCYEDVEENS"
complement (join(13598..13687,13733..13840,14157..14744,
15001..15531,15580..15808,15867..16342))
/gene="Y50E8A.6"
CDS
complement (join(13598..13687,13733..13840,14157..14744,
15001..15531,15580..15808,15867..16342))
/gene="Y50E8A.6"
/standard_name="Y50E8A.6"
/notes="contains similarity to Pfam domain: PF01224
(3AD-dependent glycerol-3-phosphate dehydrogenase),
Score=528.5, E-value=1.6e-155, N=1"
/codon_start=1
/product="Hypothetical protein Y50E8A.6"
/protein_id="CA855049.1"
/db_xref="GI:5932875"
/translation="MISRLLOKSGILAGTAATASTVAVWQRTGFSKIREAPTGSTP
GSLPTRESITIEKTRFDVAVIGGSGAGVALDAQTRGLKTAAMVEGYDCSTSS
KSKLLHGGVKYLEAKTEKDFEYHQIVQEGNERINI MKSAFFLQSFPVLVPTKYK
WQYVYMGVGVVDFLAGKNVLKPSKFKBEAIEICPTIROGFLGAMLYVDGQND
ARLVVALTAIRNGKACVNHTECIGLKDYSKNGALVKDHISETYEIHSKVVN
ATGPNDRKREADETRNKLIYVSSGHEITVAKYFCPCNAGLIVPKSSGRVIFAPW
ENVIIVGTTDDPTFESHSPVTEKEIYQILKEMNRAPEXEYBIKREDVTSIWSGIRGL
VQDPKQKENSILARGHLVDGPTGLITIAAGKLATFRMAEETVKNKLIDINKILEAQ
PCVTGMEVTEGAGNYTSMLYRTISQNYGIEQVATHLCQTYGDKVYEVKLKCKSQK
FVIGHRLHDPDFLEAEVYAVKVARIPADILARTRISLIDAAAPQVLPRIVAI
VABELKWSPEQIFVEDEKMFELKVENGSQSKKSHSGVESQLESDQNVIPQDVJD
DVMEANENKGVVSLEDOLEKLEFSDXNGCTKGRRLDHLGLNKKCESRL"
complement (join(20700..20895,21569..21768,21816..21974,
22535..22787,22833..23052,23101..23203,23647..23721))
/gene="Y50E8A.7"
CDS
complement (join(20700..20895,21569..21768,21816..21974,
22535..22787,22833..23052,23101..23203,23647..23721))
/gene="Y50E8A.7"
/standard_name="Y50E8A.7"
/codon_start=1
/product="Hypothetical protein Y50E8A.7"
/protein_id="CA850584.1"
/db_xref="GI:6425353"
/translation="NRTANNEIPFVFAQKTEIEELDAADTISHGYTVKHYVTTD
DGYTVQIQIRFVGRDRSILGSKRPVFPFMHGLFGSSHYFLNLPSSQAAYIFADAG
FDVWLGNIKRTGYSLNHTSFSTGVNFNPSLYEHSYDLRQIQIEYLDYTRHESLFY
Query Match 5.3%; Score 57.6; DB 3; Length 61864;
Best Local Similarity 48.8%; Pred. No. 0.02; Mismatches 0; Gaps 0;
Matches 155; Conservative 0
```

```
RESULT 13
AF253047
LOCUS
DEFINITION
Candida albicans DNA primase large subunit Pri2p (pri2) gene,
partial cds; large subunit Cp complex Cdc68p (cdc68) gene, complete
cds; and Cso1p (cso99) gene, partial cds.
AF253047
AF253047.1 GI:12082668
Candida albicans
Candida albicans
Saccharomycetes; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Zuckeryota; Fungi; Ascomycota; Mitosporic Saccharomycetales; Candida.
1 (bases 1 to 5230)
Suurman,E.T., Jiang,W., McCoy,M., Averett,D.R., Thompson,C.M. and
Wobbe,C.R.
Validation of Cdc68p as a novel antifungal target
Arch. Microbiol. 178 (6), 428-436 (2002)
22306560
12420162
REFERENCE 2 (bases 1 to 5230)
Suurman,E.T., Jiang,W. and Wobbe,R.
Direct Submission
Submitted (06-APR-2000) Scriptgen Pharmaceuticals, 610 Lincoln
Street, Waltham, MA 02451, USA
Location/Qualifiers
1..5230
/organism="Candida albicans"
/mol_type="genomic DNA"
/strain="1006"
/db_xref="taxon:5476"
complement (<1..>379)
/gene="pri2"
complement (<1..>379)
/gene="pri2"
/product="DNA primase large subunit Pri2p"
complement (<1..>379)
/gene="pri2"
/codon_start=1
/transl_table=12
/product="DNA primase large subunit Pri2p"
/protein_id="AAG48573.1"
/db_xref="GI:12082669"
/translation="MFRQVRRTAGRRNFDSTSTATTTTTTNGEKDPSKYLPPS
ALYASRLSFYDLPTQETILEFETWALDRKLIEESCLAFSKTLKIEFTSIKPLL
LKVLPUNPTSGGGTGVNYSKERMKD"
<834..>3992
/gene="cdc68"
<834..>3992
/gene="cdc68"
/product="large subunit Cp complex Cdc68p"
834..3992
/gene="cdc68"
/notes="Spt18p"
/codon_start=1
/transl_table=12
/product="large subunit Cp complex Cdc68p"
/protein_id="AAG48574.1"
/db_xref="GI:12082670"
/translation="MSEVVIDAGLFVKRLSIFQKLTANNI POALIIVGARSDNTYK
KSTVLQWLLGVEFHTAIYITDKKCFITSEKSKHLHLNCKPDLVELWRTKDV
ENKQFLIKLEMTTDSKYGKLDKDYDGFIDWNQILNDNNNTNDHALSAVD
LAVTVSQALAKDSEFNNTKIASNASVVMPTFVNDMMIIVDDKKITNSQTDQIE
KTEENKWKYIKLGNLQSKDFDPEYLEYCYSP1IQSGGDYDLKPSAVTDKLI
GEGVILSS1GLYKYSYCSN1ARTFLIDPTSEMETNYDFLLQLOKYLVDNLLKGVVAS
KVQTDIDVIKERRPDLVNHFTKCGWLLGMEPRDSTFILNAKTRKLTGQII SIT
IGFNLSNDKRNKDKNTKQNKQTYFALLDITIKITDDSSILLITKYSKORAA1SFS
FNDNETQKNNNNKRPGLSOTNTTGLKESNTENTAILKSKLRHENTNADANSE
KLKQETIKLHKRLOGLARSKADATDDDFPKFKKYSYVRSQIPNSYNDKEL
H1DYKQTIILPISGRPVPHINSYKSGSNGEBDFTFLINFSNPGAGNVTQKEL
PYEDSPENSLRSITIRSRRQRQMVVDYKAIQDLKXSVKREQEKKQMDVITQANLI
```

```
AF253047
5230 bp DNA linear P13 13-NOV-2002
Candida albicans DNA primase large subunit Pri2p (pri2) gene,
partial cds; large subunit Cp complex Cdc68p (cdc68) gene, complete
cds; and Cso1p (cso99) gene, partial cds.
AF253047
AF253047.1 GI:12082668
Candida albicans
Candida albicans
Saccharomycetes; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Zuckeryota; Fungi; Ascomycota; Mitosporic Saccharomycetales; Candida.
1 (bases 1 to 5230)
Suurman,E.T., Jiang,W., McCoy,M., Averett,D.R., Thompson,C.M. and
Wobbe,C.R.
Validation of Cdc68p as a novel antifungal target
Arch. Microbiol. 178 (6), 428-436 (2002)
22306560
12420162
REFERENCE 2 (bases 1 to 5230)
Suurman,E.T., Jiang,W. and Wobbe,R.
Direct Submission
Submitted (06-APR-2000) Scriptgen Pharmaceuticals, 610 Lincoln
Street, Waltham, MA 02451, USA
Location/Qualifiers
1..5230
/organism="Candida albicans"
/mol_type="genomic DNA"
/strain="1006"
/db_xref="taxon:5476"
complement (<1..>379)
/gene="pri2"
complement (<1..>379)
/gene="pri2"
/product="DNA primase large subunit Pri2p"
complement (<1..>379)
/gene="pri2"
/codon_start=1
/transl_table=12
/product="DNA primase large subunit Pri2p"
/protein_id="AAG48573.1"
/db_xref="GI:12082669"
/translation="MFRQVRRTAGRRNFDSTSTATTTTTTNGEKDPSKYLPPS
ALYASRLSFYDLPTQETILEFETWALDRKLIEESCLAFSKTLKIEFTSIKPLL
LKVLPUNPTSGGGTGVNYSKERMKD"
<834..>3992
/gene="cdc68"
<834..>3992
/gene="cdc68"
/product="large subunit Cp complex Cdc68p"
834..3992
/gene="cdc68"
/notes="Spt18p"
/codon_start=1
/transl_table=12
/product="large subunit Cp complex Cdc68p"
/protein_id="AAG48574.1"
/db_xref="GI:12082670"
/translation="MSEVVIDAGLFVKRLSIFQKLTANNI POALIIVGARSDNTYK
KSTVLQWLLGVEFHTAIYITDKKCFITSEKSKHLHLNCKPDLVELWRTKDV
ENKQFLIKLEMTTDSKYGKLDKDYDGFIDWNQILNDNNNTNDHALSAVD
LAVTVSQALAKDSEFNNTKIASNASVVMPTFVNDMMIIVDDKKITNSQTDQIE
KTEENKWKYIKLGNLQSKDFDPEYLEYCYSP1IQSGGDYDLKPSAVTDKLI
GEGVILSS1GLYKYSYCSN1ARTFLIDPTSEMETNYDFLLQLOKYLVDNLLKGVVAS
KVQTDIDVIKERRPDLVNHFTKCGWLLGMEPRDSTFILNAKTRKLTGQII SIT
IGFNLSNDKRNKDKNTKQNKQTYFALLDITIKITDDSSILLITKYSKORAA1SFS
FNDNETQKNNNNKRPGLSOTNTTGLKESNTENTAILKSKLRHENTNADANSE
KLKQETIKLHKRLOGLARSKADATDDDFPKFKKYSYVRSQIPNSYNDKEL
H1DYKQTIILPISGRPVPHINSYKSGSNGEBDFTFLINFSNPGAGNVTQKEL
PYEDSPENSLRSITIRSRRQRQMVVDYKAIQDLKXSVKREQEKKQMDVITQANLI
```



```
/codon_start=1
/product="similar to Corynebacterium glutamicum
(Brevibacterium flavum). Zn-dependent alcohol
dehydrogenases (EC 1.1.1.1)"
/db_xref="GI:26828947"
/db_xref="GI:26828947"
/translation="MEDRKTMAGVFKVGGNLEIVKMSIPKPKQGLKNEYPRVSGH
EVLEIKLGDGVDPIKFKIGHLVGWNGNCGSCRECLNNRILFCRNLIGISL
DGYASYIVPSPDALAFIPGMNABETAPILLGAVTVYCSLKNQFKGSLVGLGIG
GLGHAIOFCCKWGVYIAMSNGSKETSRLEGASHFVDIAXEGWQWQKQIGSVBC
ILLTAPPHLVQPSLQALGVNGLKLLCIIPEPPFADSLTMLLGNKSIWVYWGNSD
IQTLDPZAHKH-HKPIQTPLOAVNEVLKINEANFRSVIKIL"
complement(join(6495..7151,7309..8518,8578..8822,
8878..9034))
/note="GeneID exon scores (in order of location ranges):
37..10, 68..27, 12..46, 6..40 - GSCJ_ID dd_00586"
/codon_start=1
/product="similar to Oryza sativa pseudotsugata multicapsid
polyhedrosis virus (OpNPV). Hypothetical 29.3 kDa protein
(ORF92)"
/protein_id="AAO51532.1"
/db_xref="GI:28828948"
/translation="MNKHGPIVFLIIINRNYPVSSQNSSTELSCIFYTDSNTR
IDLSPLKAKPNSGVPHNNNNPNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNDNKKIGILVYVYKIDVYBSGLNLEASFEDXIP
NGQYEDVIFPNSGIDHCISQILFSCNDYQHKISIDQPSNVCNQIFPESYACKK
KXNLAIPTSVSSILPSISSTNSNNNTITIGENFVSLGNIQIQTIPISINCTDI
LEISGLEYCNCHIDEMVINSVDGQFYTLIENNGFIVYNSYSPSLWCPNDG
SGFYGLCNFCVDGHTSGVDSINVGFTNSIPKQLSINPEMMSTTTTITDPT
TTTTTSSNVLHGSIFSVYREVNPGIVQVRLPINTLSINDYNNSTSSFSGSS
LTPFNGCLNTQGLFYINNDNQVSNVNSFPNPZPISTSSSSSSSSSSSSSPL
LQPISSNVILLINNOIQSPENFTDFISNYSVHNLPSLPIYSPSTFYFI
SINAKGSFEPFASNLFPNPSPTPTQTLNPTQSPSTSSPTSLPTSTPTPT
SSPTGSPPTQSPSTSSPTSSPTPTPTQTLNPTQSPSTSSPTSSPTPT
VTSISGSHIPKKEIIVILTSAPGIIICVILLVFLIFKPKRKNPNFIKDYGEAGL
LDGSGNVDNNAVEBGLIIN"
complement(9434..10048)
/note="GeneID exon scores (in order of location ranges):
46..94 - GSCJ_ID dd_00587"
/codon_start=1
/product="similar to Bacillus subtilis. Hypothetical 39.0
kDa protein in ginQ-anes intergenic region"
/protein_id="AAO51533.1"
/db_xref="GI:28828949"
/translation="MNLSKVTVVPHDPKWLFLPEEKQIKVIFKDYKCKKNEXDD
TPSSAAQPLLITVNVHCSGSPVGLKAPVIDILLVANGDISELDMNERFESSY
QIRGENIAGRPPTKLPNHVWVMAFQPDNTDITQRLTFKDYLIASHIILDQVA
NLKSELASKPNSDDYWGKNLWIKCHEKSLICHWNRPNK"
complement(join(10359..10367,10573..11484,11554..11751,
11856..13705,13779..14106,14175..14618))
/note="GeneID exon scores (in order of location ranges):
-2..51, 65..70, 11..49, 142..52, 16..88, 16..56 - GSCJ_ID
dd_00588"
/codon_start=1
/product="similar to Dictyostelium discoideum (slime
mold). Protein-tyrosine phosphatase 3 (3C 3.1.3.48)
(protein-tyrosine- phosphate phosphohydrolase 3)"
/protein_id="AAL92306.2"
/db_xref="GI:28828950"
/translation="MKNIEISYLLSQOKLSHCFNSVHSYVYIKTNEILANNLLPILK
GTRDSPITLQWNDKSDNLSDPITKQENDTSDINNNNNNNNNNNNNKQCILRG
DNTNSNLKUDJILNNLNKNSDLFKLVISNLIKFKQKQKOLNIIISTFYFENF
NIRKIEILIDRSDSGSEIKLNEFESDMISIKNDQLINSKQILQFSEBEL
GSDLSNTSKFKEIVQWVNDCTVESLKNYIYKEIDQIEILLKSNQNNFN
KDLKIDSPELSKSLKSLIIEIPHYIENSTSTTASSPPASSPPSSSSSSSS
SSKLLNNLSL-LNYPNDYIQMFHKMMNESFNHYKVPNSGFVQLQSVLNHLBSKV
KEDYKGLVYNSVSDIGLFTTGHDPGIVESTYHBSIYKEYSMFSPQSOLAS
QFQYQIDQLFNOKKSTILLNSLKYIIFINCNNGNGSRNSVNSDEKLKVLKLS
FLEVPITFVNLRFSTFNNPTFDLLKKTDFLKTENILKLSINBEDQVSIQYKL
FISIALTELKRLKSLIILKNDQPADHKYSTPIKTITDTSNPKSKTSKTRFRFN
HDEITSYFKLLNLENNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNPDDSSNNNSNMYLNGOIQTDNGPFOIQLOHAEVLKQKPD
LQVSMVEQTQDKLGGDTIIGDDIDQSLSTIGQQQNSGENNYEQNRKSLINN
```

```
IIIGEVNHLIILFQNTKLIQISEKSPQIEVNEHLIYLBQYLFKOLCVSANTGVFELF
ETVRIFESLYFLKXSNQDQFLKMLFKTHINXYIILFLNMSINLIDQDDKNYKXV
TUSKMINV-SKYLTVAPHNDVNFPIYSELNGIMVHNLKMLIFGFDNDLIYAKN
ILISLITSNQNVNMAKIKGLWKEGHLVALLFKMLIILIDENLQVLIIRAKN
IVNEBFKSPFKTGLIGSVKRVSGYEVIREFPELXIZIKAYTKQOQEQOQB
QOQNEKEKEIDTITATTATATEITVISPKNLSLAPIODKGEFFPMCHQ
FOIISVLLAYKILLSVESNSEPKLIYDCLNSILNNAQIMKFEVEVLKQJHSLFLCL
ILKQNTKTLRSIQILNIPNDPNDYDWRNHYITTVQSKLTN"
complement(14950..15396)
/note="GeneID exon scores (in order of location ranges):
22..36 - GSCJ_ID dd_00578"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAO51534.1"
/db_xref="GI:28828951"
/translation="MYNKTIKSYLCFSNNNVANKSIQRTYTTNFTYTYFYHHTKAFN
DILNQKMKYQVASSPNNNGDYKMSDYRYRYMYNPKNEWHHDSFDNKPFPHTK
HNNQMTVGLKEDYBRIVEESKRTAPKPEHIKQEKMKDPSLQVD"
complement(16288..16938)
/note="GeneID exon scores (in order of location ranges):
55..75 - GSCJ_ID dd_00579"
/codon_start=1
/product="similar to Oryza sativa (japonica
cultivar-group). Hypothetical protein"
/protein_id="AAO51535.1"
/db_xref="GI:28828952"
/translation="MFCFRKPSDSIKQINDRERCPAYSNLYGTQDYATKEEYMD
PKYSGFVQIKQLGTGVECFQKAAALQKWHFLDWDVDFIKNPIAVGIVGILL
KXSNQFWILSFARINLYDQDQDSIKFGYSYGTILKDHVERGEERFVLEWVDPDGA
PDKGVAYEVLSPSEPSWLSQLGYPVTRYFQNKVVDSCNQMLKAVGNSQNVHV"
join(17987..18637,18753..19439)
/note="GeneID exon scores (in order of location ranges):
86..24, 88..94 - GSCJ_ID dd_00580"
/codon_start=1
/product="similar to Mus musculus (Mouse). 16 days embryo
head cDNA, RIKEN full-length enriched library,
clone:C130090304 product:succinate-Coenzyme A ligase,
ADP-forming, beta subunit, full insert sequence"
/protein_id="AAO51536.1"
/db_xref="GI:28828953"
Query Match 4.7%; Score 50.6; DB 3; Length 29143;
Best Local Similarity 48.8%; Pred. No. 0.69;
Matches -37; Conservative 0; Mismatches 144; Indels 0; Gaps 0;
Qy 268 AGCAGTTTCATCGTCACGCGTGAACAAATATACATTAACAATTACGGAAAAAGAGT 327
Db 2301 AGAAATTCCTGGATCAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATT 2360
Qy 328 TTAACTAAAAAGAGCTACAAATTAAGGCTATAAACAATTATGTTCAAAAGTGTTAAC 387
Db 2361 GAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATT 2420
Qy 388 TGCCCATCCGCCCTAACACTTAACCTACGCTCATTTTAACTGTAATAAAAAACGGCGTTCA 447
Db 2421 TGTTAAATCTTGAATATCGTTCCTTTTCCGACACAATAACCGTGTCTGTAATAAATTTA 2480
Qy 448 GGTGCAAGTTATATATATATATATTCCTGCTGGGCACTAAAAAATTTGCCTTTTGGTGGT 507
Db 2481 GGTGTGGTGTAAAGATCTAAACACACGCAATAAAAAAATAAAAAAATAAAAAAATAA 2540
Qy 508 ATCTGGGATCTACTCTGAAGTTAAGAGTAAAGAGCATA 548
Db 2541 TAATPACTTTCCTCAGGACCAAAAAATAAATAAATAAATAAATAAATAAATAAATAA 2581
RESULT 15
AC116979 AC116979 89549 bp DNA linear INV 12-MAR-2003
LOCUS Dictyostelium discoideum chromosome 2 map 6357117-6445670 strain
DEFINITION AX4, complete sequence.
AC116979
AC116979.2 GI:28829703
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2004, 08:09:10 ; Search time 491 Seconds
(without alignments)
9396.218 Million cell updates/sec

Title: US-09-839-894-9

Perfect score: 1086
Sequence: 1 atgaataagattttattat.....caagtagtcaaacactctag 1086

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124093041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq_29Jan04.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1990s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001as.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002s.*
- 7: Geneseqn2003as.*
- 8: Geneseqn2003bs.*
- 9: Geneseqn2003cs.*
- 10: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1086	100.0	1086	6	AAI70763
2	1086	100.0	1086	8	ABX93757 DNA encod
3	1086	100.0	7239	6	AAI70780 ETEC CS4
4	1086	100.0	7239	7	AD55748 Escherich
5	1086	100.0	7239	8	AD55774 Csa Opero
6	310.8	28.6	5336	7	AD55743 Escherich
7	298.8	27.5	5798	2	AAx84848 CS2 gene
8	298.8	27.5	5798	7	AD55745 Escherich
9	50	4.6	2000	7	ADA71938 Rice gene
10	49.4	4.5	10279	6	ABL33591 Human imm
11	49.4	4.5	10279	6	ABL92277 Chemicall
12	49.4	4.5	10279	6	AD22328 Chemicall
13	48.2	4.4	9504	4	ABK28407 DNA trans
14	47.2	4.3	5139	3	AAA70139 Plasmogiu
15	45.2	4.2	4176	4	ABL10866 Drosophil
16	44.8	4.1	13133	6	ABK31231 Signal tr
17	44	4.1	2000	7	ADA71938 Rice gene
18	43.8	4.0	6971	6	ABN80061 Human che
19	43.4	4.0	2036	4	ABL01847 Drosophil
20	43.4	4.0	6080	4	ABL01846 Escherich
21	43.2	4.0	50000	6	ABL55643 AmbPV gen
22	42.8	3.9	4985	3	ABQ75107 Anopheles
23	42.8	3.9	4985	9	ACF79720 Mosequit

C 24	42.8	3.9	6137	6	ABL34450 Human met
C 25	42.8	3.9	6137	6	ABL70127 Chemicall
C 26	42.6	3.9	641	6	ABQ56694 Human col
C 27	42.4	3.9	19734	6	ABL33933 Human imm
C 28	42.2	3.9	9095	6	ABK28448 DNA trans
C 29	41.6	3.8	6209	4	AAx284751 Genomic s
C 30	41.6	3.8	15548	6	ABL34155 Human imm
C 31	41.6	3.8	110000	6	ABN90521_10
C 32	41.2	3.8	454	7	ABF40601 Toxicity
C 33	41.2	3.8	454	9	ABD55668 Toxicity-
C 34	41.2	3.8	16228	6	ABL70459 Chemicall
C 35	41.2	3.8	16228	6	ABK28448 DNA trans
C 36	41.2	3.8	16228	6	ABK28448 DNA trans
C 37	41.2	3.8	16228	6	ABK28448 DNA trans
C 38	41.2	3.8	16228	6	ABK28448 DNA trans
C 39	41.2	3.8	16228	6	ABK28448 DNA trans
C 40	41.2	3.8	16228	6	ABK28448 DNA trans
C 41	41.2	3.8	16228	6	ABK28448 DNA trans
C 42	41.2	3.8	16228	6	ABK28448 DNA trans
C 43	40.8	3.8	6534	6	ABL32470 Human imm
C 44	40.8	3.8	7833	9	ABD54183 Pretreate
C 45	40.8	3.8	7833	9	ABD54183 Pretreate

ALIGNMENTS

RESULT 1
AAI70763
ID AAI70763 standard; DNA; 1086 BP.
XX AAI70763;
AC AAI70763;
XX
DT 18-FEB-2002 (first entry)
XX
DE ETEC CS4 pilus csae gene coding region.
XX
KW CS4 pilus; enterotoxigenic; ETEC; csae operon; csae gene; fimbrial;
KW vaccine; diarrhoea; antibacterial; antidiarrheic; cs.
XX
OS Escherichia coli.
XX
EH Key Location/Qualifiers
FT s-g_peptide i..69
FT /*tag= a
FT mat_peptide 70..1083
FT /*tag= b
XX
XX WO200181582-A2.
XX
XX 01-NOV-2001.
XX
XX 20-APR-2001; 2001WO-US012914.
XX
XX 20-APR-2000; 2000US-0198686P.
XX
XX (JYMA-) UNIV MARYLAND BALTIMORE.
XX
XX Altboun Z, Levine MM, Barry EM;
XX
XX WPI; 2002-049280/06.
XX
XX P-PSDB; AAM50343.
XX
XX New nucleotide sequence, useful as immunogenic agent for generating
XX immune response against recombinant product of the operon, comprises csae
XX operon which encodes enterotoxigenic Escherichia coli-CS4 pilus.
XX
XX Claim 31; Page 57-58; 81pp; English.
XX
XX The present sequence is that of the coding region of the csae gene of
XX enterotoxigenic Escherichia coli (ETEC) strain E11881A. The csae operon
XX (see AAI70780) was isolated from a genomic library of this strain.
XX
XX Sequencing revealed 5 contiguous genes, csae-csaE (see AAI70759-63),

CC flanked by 2 insertion elements. The *csa* operon encodes the synthesis of
 CC ETEC-CS4 pili, and has been expressed in attenuated Shigella strain
 CC CVD1204 guABA, constructing the Shigella expressing CS4 fimbriae vaccine
 CC strain CVD1204 (pGA2-CS4). The *csaE* gene encodes a 40 kDa tip associated
 CC protein (CsaD, see AM50341) that shows homology with similar proteins
 CC from other ETEC fimbriae. The *csa* operon, and the *csaA-csaE* genes, are
 CC useful in the production of recombinant CsaA-CsaE polypeptides that are
 CC used in claimed immunogenic compositions to prevent ETEC colonisation,
 CC and hence to protect against diarrhoea
 XX
 SQ Sequence 1086 BP; 366 A; 187 C; 202 G; 331 T; 0 U; 0 Other;
 Query March 100.0%; Score 1086; DB 6; Length 1086;
 Best Local Similarity 100.0%; Pred No. 6 5e-254;
 Matches 1086; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGAATTAAGATTTTAAATTTTACATTTGTTTCTTCTCAGTACTTTTACATTTGCT 60
 DB 1 ATGAATTAAGATTTTAAATTTTACATTTGTTTCTTCTCAGTACTTTTACATTTGCT 60
 QY 61 GTATCGGAGATAAAATCCCGGAGATCAAGCAATACTAATATTTTGGCCGCGTGAC 120
 DB 61 GTATCGGAGATAAAATCCCGGAGATCAAGCAATACTAATATTTTGGCCGCGTGAC 120
 QY 121 AGGAACGAATCTCCCCCAACATAATATTAATAACCATATTACAGCATACAGTGAA 180
 DB 121 AGGAACGAATCTCCCCCAACATAATATTAATAACCATATTACAGCATACAGTGAA 180
 QY 181 AGTCATACTCTGTATGATAGATGACATTTTATATGTTTGTCTTCTCACAATACACTTAAT 240
 DB 181 AGTCATACTCTGTATGATAGATGACATTTTATGTTTGTCTTCTCACAATACACTTAAT 240
 QY 241 GGACATGTCACACAGTCAGAACTCCTAGCAGTTCATCGGTCAGCGGTGAACAAATATA 300
 DB 241 GGACATGTCACACAGTCAGAACTCCTAGCAGTTCATCGGTCAGCGGTGAACAAATATA 300
 QY 301 ACATTACAATTTACGAAAAAAGAAAGTTTAAATAAAGAGAGCTACAAATTAAGGCTAT 360
 DB 301 ACATTACAATTTACGAAAAAAGAAAGTTTAAATAAAGAGAGCTACAAATTAAGGCTAT 360
 QY 361 AAACAATTTATGTTTCAAAGTGTAACTGCGCCATCCCGCCTTAACACTTAACCTCAGCTCAT 420
 DB 361 AAACAATTTATGTTTCAAAGTGTAACTGCGCCATCCCGCCTTAACACTTAACCTCAGCTCAT 420
 QY 421 TTTAACTGTAATAAAACGGGCTTCAAGTCAAGTTCATATATTTATTTATTTCTCTGGC 480
 DB 421 TTTAACTGTAATAAAACGGGCTTCAAGTCAAGTTCATATATTTATTTATTTCTCTGGC 480
 QY 481 GAACTAAAAAATTTGCGCTTTTGGTATCTGGGATGCTACTCTGGAAGTAAAGATAAAA 540
 DB 481 GAACTAAAAAATTTGCGCTTTTGGTATCTGGGATGCTACTCTGGAAGTAAAGATAAAA 540
 QY 541 AGACGATATAGTGAGACCTATGGAACTTACACTATAATAATACATTTAAATTAACGTAT 600
 DB 541 AGACGATATAGTGAGACCTATGGAACTTACACTATAATAATACATTTAAATTAACGTAT 600
 QY 601 AAGGGAATATTCAGATATGTTTCACTCAGTTCAAAAGTGAGCTCGGTCGATCTTAAC 660
 DB 601 AAGGGAATATTCAGATATGTTTCACTCAGTTCAAAAGTGAGCTCGGTCGATCTTAAC 660
 QY 661 TTGCTTCAAACTGTTGGGGGACATATATTTGGAAGAAATTTCTTGATATGCTTTTAT 720
 DB 661 TTGCTTCAAACTGTTGGGGGACATATATTTGGAAGAAATTTCTTGATATGCTTTTAT 720
 QY 721 GATGATATAGTAAACAGAGCTCTTTGGAGATAAGATTTTCAGGATAACAAATCTTAA 780
 DB 721 GATGATATAGTAAACAGAGCTCTTTGGAGATAAGATTTTCAGGATAACAAATCTTAA 780
 QY 781 TCTGATGGAAATTTTATCTAAGGAATAAATATGATGACACCAAGAAATTCATATACT 840
 DB 781 TCTGATGGAAATTTTATCTAAGGAATAAATATGATGACACCAAGAAATTCATATACT 840
 QY 841 TTGTGAC-TCTCTTGGCGGTAAAAAGTTTAACTCCACAAATGGAACGTCATTAATAT 900

DB 841 TTGTCACTTCTCTTGGCGGTAAAGCTTTAACTCCAACTGGAACGTCAATAATAT 900
 QY 901 GCTGACGAGCTTCTCTGGAACCACTGGAATAGATTACAGCTGTCAACCATGCCAGAA 960
 DB 901 GCTGACGAGCTTCTCTGGAACCACTGGAATAGATTACAGCTGTCAACCATGCCAGAA 960
 QY 961 ATCAGATTTCCGGTGTGTTGGCTTGGCTTGGATGCAATTTGATGCAAAAGTGGAAAT 1020
 DB 961 ATCAGATTTCCGGTGTGTTGGCTTGGCTTGGATGCAATTTGATGCAAAAGTGGAAAT 1020
 QY 1021 CCGAGGCTGGACAATATATATGAGTAATTAATGTTACTTTTCACACCAAGTACTCAACA 1080
 DB 1021 CCGAGGCTGGACAATATATATGAGTAATTAATGTTACTTTTCACACCAAGTACTCAACA 1080
 QY 1081 CTCTAG 1086
 DB 1081 CTCTAG 1086
 RESULT 2
 ABX93757
 ID ABX93757 standard; DNR; 1086 BP.
 XX
 XX ABX93757;
 XX AC
 XX 16-JUN-2003 (first entry)
 XX
 DE DNA encoding *csa* operon recombinant product, CsaE.
 XX
 XX CsaE; gene; ds; *csa* operon; CS4 pilus; bacterial pili protein; ETEC; CS4;
 KW enterotoxigenic Escherichia coli; multivalent Shigella-ETEC; diarrhoea;
 KW Shigella; antibacterial.
 XX
 XX Escherichia coli.
 XX
 FH Key Location/Qualifiers
 CDS 1..1086
 FT /*tag= a
 FT /product= "CsaE"
 XX
 XX US2002:76868-A1.
 XX
 PD 28-NOV-2002.
 XX
 XX 20-APR-2001; 2001US-00839894.
 XX
 XX 20-APR-2000; 2000US-0198626P.
 PR
 XX (ALTB/) ALTBOM Z.
 PA (LEVI/) LEVINE M M.
 PA (BARR/) BARRY E M.
 XX
 XX Altbom Z, Levine M, Barry E;
 FI
 XX WPI; 2003-352604/33.
 DR P-PSDB; ABU09014.
 XX
 DR Novel isolated nucleic acid comprising *csa* operon encoding proteins
 PT required for producing CS4 pili, useful for generating immune response in
 PT vertebrate against the enterotoxigenic Escherichia coli.
 XX
 PS Claim 31; Page 30-31; 58pp; English.
 XX
 CC The invention relates to an isolated nucleotide sequence comprising a *csa*
 CC operon (encoding proteins required for producing CS4 pili) or its
 CC functional fragment. An immunogenic composition comprising a recombinant
 CC product of a *csa* operon and a carrier, is useful for generating an immune
 CC response in a subject, which involves contacting the subject with the
 CC immunogenic composition. The recombinant product of the composition is
 CC the CS4 antigen (bacterial pili protein) and is provided in an acellular
 CC or cellular composition. The nucleic acid is useful for producing a
 CC polypeptide product from a *csa* operon or functional fragment, which

CC involves providing a nucleic acid, introducing the csa operon in an
CC expression vector, such that a recombinant host cell is produced and
CC subjecting the recombinant host cell to conditions such that a protein
CC from the csa operon is expressed. The nucleic acid encoding at least an
CC immunogenic portion of the csa operon or a polypeptide encoded by the
CC nucleic acid is useful for generating an immune response in a vertebrate
CC against ETEC (enterotoxigenic Escherichia coli). The nucleic acid is also
CC useful for identifying polynucleotides encoding other proteins with
CC biological functions similar to that of the csa operon and for creating a
CC multivalent Shigella-EPEC immunogenic composition that will protect from
CC diarrhoea caused by either Shigella or EPEC expressing ETEC strains. This
CC sequence represents DNA encoding a recombinant product of a csa operon,
CC CsaE
XX
XX Sequence 1086 BP; 366 A; 187 C; 202 G; 331 T; 0 U; 0 Other;
Query Match 100.0%; Score 1086; DB 8; Length 1086;
Best Local Similarity 100.0%; Pred. No. 6.5e-254;
Matches 1086; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGAATAAGATTATTTATTTTATTTTACATTTGTTTCTCTCTCAGTACTTTTACATTTGCT 60
Db |||||||
QY 1 ATGAATAAGATTATTTATTTTATTTTACATTTGTTTCTCTCTCAGTACTTTTACATTTGCT 60
Db |||||||
QY 61 GTATCGCGAGATAAATTCGCGAGATGAAGCATTAAGTAACTTAATTTTGGCCCGCGTGAC 120
Db |||||||
QY 61 GTATCGCGAGATAAATTCGCGAGATGAAGCATTAAGTAACTTAATTTTGGCCCGCGTGAC 120
Db |||||||
QY 121 AGCAACGATCTTCCGCCAACATATATATTAATTAACCATATTTACAGCATACAGTGAA 180
Db |||||||
QY 121 AGCAACGATCTTCCGCCAACATATATATTAATTAACCATATTTACAGCATACAGTGAA 180
Db |||||||
QY 181 AGTCATCTCTGTATGATAGTACATCTTTTATTTGTTTGTCTTCTCACATACACTTAAT 240
Db |||||||
QY 181 AGTCATCTCTGTATGATAGTACATCTTTTATTTGTTTGTCTTCTCACATACACTTAAT 240
Db |||||||
QY 241 GGAGCATGTCCAAACAGTGAATCCTAGAGTTCATCGGTTCAGCGGTGAACCAATATA 300
Db |||||||
QY 241 GGAGCATGTCCAAACAGTGAATCCTAGAGTTCATCGGTTCAGCGGTGAACCAATATA 300
Db |||||||
QY 301 ACATTACAAATTCAGGAAAAGAGTTTAAATAAAGAGAGCTTACAAATTAAGGCTAT 360
Db |||||||
QY 301 ACATTACAAATTCAGGAAAAGAGTTTAAATAAAGAGAGCTTACAAATTAAGGCTAT 360
Db |||||||
QY 361 AAACAAATTTGTTTCAAAAGTGAATTAAGTCCATCGGCTTAACTTAACCTCAGCTCAT 420
Db |||||||
QY 361 AAACAAATTTGTTTCAAAAGTGAATTAAGTCCATCGGCTTAACTTAACCTCAGCTCAT 420
Db |||||||
QY 421 TTAACTGTATAAATAAAGCGGCTTCAGTGCAGTTTATTTATATATTTATTTTCCCTGGC 480
Db |||||||
QY 421 TTAACTGTATAAATAAAGCGGCTTCAGTGCAGTTTATTTATATATTTATTTTCCCTGGC 480
Db |||||||
QY 481 GAACATAAATAATTTGCCCTTTTGGTGTATCTGGATGCTACTCTGAAGTTAAGAGTAAAA 540
Db |||||||
QY 481 GAACATAAATAATTTGCCCTTTTGGTGTATCTGGATGCTACTCTGAAGTTAAGAGTAAAA 540
Db |||||||
QY 541 AGACGATATAGTGAGACCTATGGAACCTTACACTATAAATACACTATTAAATTAACCTGAT 600
Db |||||||
QY 541 AGACGATATAGTGAGACCTATGGAACCTTACACTATAAATACACTATTAAATTAACCTGAT 600
Db |||||||
QY 601 AAGGGAATAATTCAGATATGTTTACTCAGTTCAAAAGTGAGCTCCGCTGCACTTAAC 660
Db |||||||
QY 601 AAGGGAATAATTCAGATATGTTTACTCAGTTCAAAAGTGAGCTCCGCTGCACTTAAC 660
Db |||||||
QY 661 TTGCGTCCAACTGGTGGGGCACATATATGGAAGAAATTCGTTGATATGCTTTTAT 720
Db |||||||
QY 661 TTGCGTCCAACTGGTGGGGCACATATATGGAAGAAATTCGTTGATATGCTTTTAT 720
Db |||||||
QY 721 GATGATATAGTACTAACAAGAGCTCTTTGGAGATAAGATTTTACAGATAACAATCCTAAA 780
Db |||||||
QY 721 GATGATATAGTACTAACAAGAGCTCTTTGGAGATAAGATTTTACAGATAACAATCCTAAA 780
Db |||||||
QY 781 TCTGATGGGAATTTTATCTAAGGAAAATAAATGATGACACCAAGAAATTCATATACT 840
Db |||||||

Db 781 TCTGATGGGAATTTTATCTAAGGAAAATAAATGATGACACCAAGAAATTCATATACT 840
QY 841 TTGTCACTCTCTCTGGCGGTAAAGATTTAACCTCAACAAATGGAACGCTCATTAATTT 900
Db |||||||
QY 841 TTGTCACTCTCTCTGGCGGTAAAGATTTAACCTCAACAAATGGAACGCTCATTAATTT 900
Db |||||||
QY 901 GCTGACGAGCTCTCTCTGGAAACAACTGGAAATAGAAATTACAGCTGTACCAATGCCAGAA 960
Db |||||||
QY 901 GCTGACGAGCTCTCTCTGGAAACAACTGGAAATAGAAATTACAGCTGTACCAATGCCAGAA 960
Db |||||||
QY 961 ATCACTGTTTCCGCTGTTTGGCTGGACCTTTGCAATTCGATGCAAAAAGTGGAAAT 1020
Db |||||||
QY 961 ATCACTGTTTCCGCTGTTTGGCTGGACCTTTGCAATTCGATGCAAAAAGTGGAAAT 1020
Db |||||||
QY 1021 CCCGAGCTGGACAAATATATGGGTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 1080
Db |||||||
QY 1021 CCCGAGCTGGACAAATATATGGGTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 1080
Db |||||||
QY 1081 CTCCTAG 1086
Db |||||||
QY 1081 CTCCTAG 1086
Db |||||||
RESULT 3
AAI70780
ID AAI70780 standard; DNA; 7239 BP.
XX
AC AAI70780;
XX
DT 18-FEB-2002 (first entry)
XX
DE ETEC C54 pilus csa operon.
XX
KW C54 pilus; enterotoxigenic; ETEC; csa operon; csaA gene; csaB gene;
KW csaC gene; csaD gene; csaE gene; fimbrial; vaccine; diarrhoea;
KW antibacterial; antidiarrheic; ds.
XX
OS Escherichia coli.
XX
Key Location/Qualifiers
CDS 283..999
FT /*tag= a
FT /note= "csaA (claim 15), encodes AAM50339"
FT sig_peptide 283..342
FT /*tag= b
FT mat_peptide 343..996
FT /*tag= c
FT CDS 1028..1531
FT /*tag= d
FT /note= "csaB (claim 22), encodes AAM50340"
FT sig_peptide 1028..1096
FT /*tag= e
FT mat_peptide 1097..1528
FT /*tag= f
FT CDS 1589..4192
FT /*tag= g
FT /note= "csaC (claim 25), encodes AAM50341"
FT sig_peptide 1589..1654
FT /*tag= h
FT mat_peptide 1655..4189
FT /*tag= i
FT CDS 4196..5281
FT /*tag= j
FT /note= "csaE (claim 31), encodes AAM50343"
FT sig_peptide 4196..4264
FT /*tag= k
FT mat_peptide 4265..5278
FT /*tag= l
FT CDS complement (5790..6119)
FT /*tag= m
FT /note= "(claim 28) csaD, encodes AAM50342"
XX


```
SQ Sequence 5798 BP; 1904 A; 874 C; 1221 G; 1799 T; 0 U; 0 Other;

Query Match      27.5%; Score 298.8; DB 2; Length 5798;
Best Local Similarity 58.5%; Pred. No. 1.9e-62;
Matches 566; Conservative 0; Mismatches 387; Indels 15; Gaps 2;

QY 127 GAATCTTCCCCAAACATAATATATTAATAACCAATATACAGATACAGTGAAGTCAAT 186
D5 4583 GCATCGTATCCTGCTCATTTATATATTTATGATGATGCTGGTTACAAATAAAGATCAC 4642

QY 187 ACTCTGTATGATAGGATGACTTTTATGTTTCTCTCTCAACAATACACTTAATGAGCA 246
D5 4643 TCTCTTTTTCACAGGATGAGCTTTTATGTTTCTCTCTCAACAATACACTTAATGAGCA 4702

QY 247 TGTCCAAACAGTCAGAAATCCTAGCAGTTCATCGCTCAGCGGTGAAACAAATATAACATTA 306
D5 4703 TGTCCGACAGGAGAAACCTCCA-----AATCTCTCAAGGGAGACTAATATTAGCTA 4756

QY 307 CAATTTACGGAAAAAGAGATTTTAATAAAGAGAGCTACAAATTAAGGCTATAAACAA 366
D5 4757 AATTTTACTGAAAGAAAGAGTCTGGCCAGAGAAACATTAACCTAAAGGATATAAGAGA 4816

QY 367 TTATCTTCAAAAGTGTAACTGC-----CCATCCGGCTACACCTTAACCTCAGCT 417
D5 4817 TTTTATATGAATCAGATAGATGCAATTCATATGTCGATAAATGAATCTCAATTTCTCAT 4876

QY 418 CATTTTAACTGTAAATAAAACGGGCTTCAGGTGCAAGTTTATATTTATATATATCTCTGCT 477
D5 4877 ACTGTTAAATGTAGTTCAITTCACAGAGGAGTAGAATTCACITTAATATATCCACAA 4936

QY 478 GCGCAATCAAAAATTTGCTTTTGTGTGATCTGGGATGCTACTCTGGAAGTTAAGATA 537
D5 4937 GGTGAAATTCATGGCTTCTAATCTGAGGTATATGGGAGGCAACACTAGAGTTACGAGTC 4996

QY 538 AAAGACGATATAGTGAGACTATGAACTTACACTATAATATCTACTATTAATTAACCT 597
D5 4997 AAAAGGCATTCACGACTATAATCATGTGTACTTACAAAGTTAATATCACAGTTGATTTGACA 5056

QY 598 GATAAGGGAATATTCAGATATGTTTACCTCAGTTCAAAAGTGCAGCTCCGCTGATCTTT 657
D5 5057 GACAAAGGAATATTCAGTCTGACACCAAGTTTCATAGCGATCTTAGAATGTATCTG 5116

QY 658 AACTTCGGTCCAACTGGTGGGGGACATATATTTGGAAGAAATTTCTGTGATATGTGTTTT 717
D5 5117 AATTTACGTCCTCAAGGTAATGTTAAATATTTCTGTTAGTAACTGCTTGAGATGTGCTC 5176

QY 718 TATCAGGATATAGTACACGAGCTCTTTGCGATATAGATTCAGGATACCAATCCT 777
D5 5177 TATGATGGCTATAGTACACATAGTCAAGTATAGAAATGAGGTTTCAGGATGACTCAAA 5236

QY 778 AAATCTGATGGGAAATTTTATCTAAGGAAATTAATGATGACACCAAGAAATTCATAT 837
D5 5237 ACAGGAATATGAATATATCTTATATAAATCTGAGAGCCATTAATAAATTTGCCATAT 5296

QY 838 ACTTTGTCACTTCTCTGGCGGGTAAAGTTTAACTCCACAAATGGAAGCTCATTAAT 897
D5 5297 AAATCTTCTCTCTTTAGGAGGACGAGATTTTATCCAAATATGAGAGGCTTTTACT 5356

QY 898 ATTCTCAGCAGGCTTCTCTGGAACAACTGGAATAGATTAACAGTGTACATGCCA 957
D5 5357 ATTAATGATATCTCTGTCATTTTATATAAGTGAATCGTATTAAGTCTGTATCTCTACCA 5416

QY 958 GAAATCAGTGTCCGGTGTCTGTGTGGCTGCGAGTTTGCAATTTGGATGCAAAAGTGGAA 1017
D5 5417 CAGATTAGTATTCAGTACTATGCTGCGCCAGCAACTTGACATTTATGTGACAGCTAAT 5476

QY 1018 AATCCGAGGCTGGAACATATATCGGTAATATTAATGTTACTTTCAACCAAGTAGTCAA 1077
D5 5477 AATCCAGAGCGGCTGAGTATTCAGGATATTAACGTAACATTTACTCTCTAGTAGTTCA 5536

QY 1078 ACATCTTA 1085
D5 5537 AGTCTGTA 5544
```

RESULT 8

```
AD55745
ID AAD55745 standard; DNA; 5798 BP.
XX
AC AAD55745;
XX
DT 07-AUG-2003 (first entry)
XX
DE Escherichia coli CS2 operon.
XX
KW Coli surface antigen; CS; vaccine; diarrhoea; bacterial infection;
KW antigen; cotB; cotA; cotC; gene; ds.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT misc_feature 499..1215
FT /tag= a
FT /note= "cotB gene"
FT misc_feature 1235..1767
FT /tag= b
FT /note= "cotA gene"
FT misc_feature 1836..4436
FT /tag= c
FT /note= "cotC gene"
FT misc_feature 4451..5545
FT /tag= d
FT /note= "cotD gene"
XX
WO2003022306-A2.
XX
PD 20-MAR-2003.
XX
PF 11-SEP-2002; 2002NO-GB004123.
XX
PR 11-SEP-2001; 2001GB-00021998.
XX
(ACAM-) ACAMBI5 RES LTD.
XX
PI Turner AK, Greenwood J, Stephens JC, Beavis JC, Darsley MJ;
XX
WPI; 2003-301009/29.
XX
PT New bacterial cell expressing three or more coli surface antigens, useful
PT for manufacturing a medicament, i.e. a vaccine, for vaccination against
PT diarrhoea.
XX
PS Disclosure; Page 99-101; 58pp; English.
XX
CC The invention relates to a bacterial cell expressing three or more coli
CC surface (CS) antigens. The bacterial cell is useful for manufacturing a
CC medicament, i.e. a vaccine, for vaccination against diarrhoea. The
CC vaccine is also useful for targeting bacterial infection. The present
CC sequence is Escherichia coli CS2 operon containing cotB, cotA, cotC and
CC cotD gene. This sequence is used in the invention
XX
SQ Sequence 5798 BP; 1904 A; 874 C; 1221 G; 1799 T; 0 U; 0 Other;

Query Match      27.5%; Score 298.8; DB 7; Length 5798;
Best Local Similarity 58.5%; Pred. No. 1.9e-62;
Matches 566; Conservative 0; Mismatches 387; Indels 15; Gaps 2;

QY 127 GAATCTTCCCCAAACATAATATATTAATAACCAATATACAGATACAGTGAAGTCAAT 186
D5 4583 GCATCGTATCCTGCTCATTTATATTTATGATGATGCTGGTTACAAATAAAGATCAC 4642

QY 187 ACTCTGTATGATAGGATGACTTTTATGTTTCTCTCTCAACAATACACTTAATGAGCA 246
D5 4643 TCTCTTTTTCACAGGATGAGCTTTTATGTTTCTCTCTCAACAATACACTTAATGAGCA 4702

QY 247 TGTCCAAACAGTCAGAAATCCTAGCAGTTCATCGCTCAGCGGTGAAACAAATATAACATTA 306
D5 4703 TGTCCGACAGGAGAAACCTCCA-----AATCTCTCAAGGGAGACTAATATTAGCTA 4756

QY 307 CAATTTACGGAAAAAGAGATTTTAATAAAGAGAGCTACAAATTAAGGCTATAAACAA 366
D5 4757 AATTTTACTGAAAGAAAGAGTCTGGCCAGAGAAACATTAACCTAAAGGATATAAGAGA 4816

QY 367 TTATCTTCAAAAGTGTAACTGC-----CCATCCGGCTACACCTTAACCTCAGCT 417
D5 4817 TTTTATATGAATCAGATAGATGCAATTCATATGTCGATAAATGAATCTCAATTTCTCAT 4876

QY 418 CATTTTAACTGTAAATAAAACGGGCTTCAGGTGCAAGTTTATATTTATATATCTCTGCT 477
D5 4877 ACTGTTAAATGTAGTTCAITTCACAGAGGAGTAGAATTCACITTAATATATCCACAA 4936

QY 478 GCGCAATCAAAAATTTGCTTTTGTGTGATCTGGGATGCTACTCTGGAAGTTAAGATA 537
D5 4937 GGTGAAATTCATGGCTTCTAATCTGAGGTATATGGGAGGCAACACTAGAGTTACGAGTC 4996

QY 538 AAAGACGATATAGTGAGACTATGAACTTACACTATAATATCTACTATTAATTAACCT 597
D5 4997 AAAAGGCATTCACGACTATAATCATGTGTACTTACAAAGTTAATATCACAGTTGATTTGACA 5056

QY 598 GATAAGGGAATATTCAGATATGTTTACCTCAGTTCAAAAGTGCAGCTCCGCTGATCTTT 657
D5 5057 GACAAAGGAATATTCAGTCTGACACCAAGTTTCATAGCGATCTTAGAATGTATCTG 5116

QY 658 AACTTCGGTCCAACTGGTGGGGGACATATATTTGGAAGAAATTTCTGTGATATGTGTTTT 717
D5 5117 AATTTACGTCCTCAAGGTAATGTTAAATATTTCTGTTAGTAACTGCTTGAGATGTGCTC 5176

QY 718 TATCAGGATATAGTACACGAGCTCTTTGCGATATAGATTCAGGATACCAATCCT 777
D5 5177 TATGATGGCTATAGTACACATAGTCAAGTATAGAAATGAGGTTTCAGGATGACTCAAA 5236

QY 778 AAATCTGATGGGAAATTTTATCTAAGGAAATTAATGATGACACCAAGAAATTCATAT 837
D5 5237 ACAGGAATATGAATATATCTTATATAAATCTGAGAGCCATTAATAAATTTGCCATAT 5296

QY 838 ACTTTGTCACTTCTCTGGCGGGTAAAGTTTAACTCCACAAATGGAAGCTCATTAAT 897
D5 5297 AAATCTTCTCTCTTTAGGAGGACGAGATTTTATCCAAATATGAGAGGCTTTTACT 5356

QY 898 ATTCTCAGCAGGCTTCTCTGGAACAACTGGAATAGATTAACAGTGTACATGCCA 957
D5 5357 ATTAATGATATCTCTGTCATTTTATATAAGTGAATCGTATTAAGTCTGTATCTCTACCA 5416

QY 958 GAAATCAGTGTCCGGTGTCTGTGTGGCTGCGAGTTTGCAATTTGGATGCAAAAGTGGAA 1017
D5 5417 CAGATTAGTATTCAGTACTATGCTGCGCCAGCAACTTGACATTTATGTGACAGCTAAT 5476

QY 1018 AATCCGAGGCTGGAACATATATCGGTAATATTAATGTTACTTTCAACCAAGTAGTCAA 1077
D5 5477 AATCCAGAGCGGCTGAGTATTCAGGATATTAACGTAACATTTACTCTCTAGTAGTTCA 5536

QY 1078 ACATCTTA 1085
D5 5537 AGTCTGTA 5544
```


Db 4703 TGTCCGACGAGGAAAACCTCC-----AATCTCTCAAGGGGAGACTAAATATTAGCTA 4756
Qy 307 CAATTTACGGAAGAAAGATTATATAAAGAGAGCTACAAATTAAGGCTATATAACAA 366
Db 4757 ATATTTACTGAAAGAAAGAGTCTGCCGAGAAACATTAACCTTAAAGGATATAAGAGA 4816
Qy 367 TTAATTGTTCAAAGTGTAACTGC-----CATCGGGCTAACACTTAACCTCAGCT 417
Db 4817 TTTTATATGAATCAGATAGATGCAATTCATATGTCGATGATGATGATGATGATGATGAT 4876
Qy 418 CATTTTAACTGTAATAAAGACCGGCTTCAGGTGCAAGTTTATATATATATATATCTGCT 477
Db 4877 ACTGTTAAATGTTAGTTCATTCACAGAGAGTAGATTTCACITATATATATCCACAA 4936
Qy 478 GCGCAACTAAAAAATTTGCTTTTGGTGATCTCTGGATGCTACTCTGAGTAAAGATTAAGATTA 537
Db 4937 GGTGAATTTGATGGGCTTCTAAGTGGATATATGGAGGCAACACTAGATTAGCAGTC 4996
Qy 538 AAAAGACGATATAGTACAGCTTATGAACTTACACTTATAAATATCACTATTAATTAATTA 597
Db 4997 AAAAGGCATTTAGCACTATATATCATGCTACTTACAAAGTTAATATACAGCTGATTTGACA 5056
Qy 598 GATAAGGAAATATTCAGATATGTTTACCTCAGTTTCAAAAAGTACGCTCGCTCGATCTT 657
Db 5057 GACAAAGGAATATTCAGTCTGGACACCAAGTTTCATAGCATCTTGAATTAATGATCTG 5116
Qy 658 AACTTTCGCTCCAACTGCTGGGGGACATATATTTGGAGAAATTTCTGATATGCTGCTT 717
Db 5117 AATTTACGTTCTTCTTTTAGGAGGACGAGAGTTTATCCAAATTAATGGAGAGGCTTTTACT 5356
Qy 898 AITGCTGACGAGCTTCTCTGAAACAAACTGAAATAGAAATACACTCTCACCAGTCCA 957
Db 5357 ATTAATGATACCTCGCACTTGTATTAACTGGAATCGTATTAAGTCTGTATCCCTTACCA 5416
Qy 958 GAAATCAGTGTTCGGGTGTTGTTGGCTCGACGTTTGCAATGGATGCAAAAGTGGAA 1017
Db 5417 CAGATTAGTATTCAGTACTATGCTGGCCAGCAACTTGACATTTATGTCAGAGCTAAAT 5476
Qy 1018 AATCCGAGGCTGGACAAATATATGAGGTAAATATTAATGTTACTTTTACACCAAGTACTCAA 1077
Db 5477 AATCCAGAGCGGCTGAGTATTTCAGGATACCTTAACGTAACATTTTACTCTTAGTACTTCA 5536
Qy 1078 ACACCTCTA 1085
Db 5537 AGTCTGTA 5544

RESULT 9

ADA71938/c
ID ADA71938 standard; DNA; 2000 BP.
XX
AC ADA71938;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 5263.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX

Osryza sativa.
XX WO2003000898-A1.
XX 03-JAN-2003.
XX 22-JUN-2001; 2001WO-IB001105.
XX 22-JUN-2001; 2001WO-IB001105.
PA (SYN) SYNGENTA PARTICIPATIONS AG.
XX Chang H. Chen W. Cooper B. Glazebrook J. Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX WPI; 2003-175290/17.
XX
XX Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
PS Claim 27; SEQ ID NO 5263; 899pp; English.
XX
XX The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC the expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
XX illustrate the invention.
SQ Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match 4.6%; Score 50; DB 7; Length 2000;
Best Local Similarity 9.0%; Pred. No. 0.04;
Matches 51; Conservative 318; Mismatches 298; Indels 3; Gaps 2;
Qy 221 CTCTCACAATACACTTAATGGAGCATGTCACCAACAGTCCAGAAATCTAGAGTTCTCGG 280
Db 1019 CTGKWTYFRSNTWYAMKKYTKYTAYSSTWKYWAYKRAYAWRSRKTWCTGCKG 960
Qy 281 TCAGCGTGAACAATAATAACATTACATTTACGGAATAAGAAATTTAATAAAGAG 340
Db 959 MATYCTCTKWAAGRWWRWAWCWYCCMKWKTSCMWWKYRTWCWYTMWGAAYAY 900
Qy 341 AGTCAAAATTAAGGCTATAAACAATTTATGTTCAAAAGTGTTAACCTGCCATCCGCGC 400
Db 899 YAMRRRTYKWSRWRYWTKWATWTKCMKAWYATGATWATWRYTYTYCYAM 840
Qy 401 TTAACACTTAACAGCTCATTTTAACTGTAAATAAAGCGG-CCTTCAGGTGCAAGTTTA 459
Db 839 TCAKCKYKXMTKWTWACARATSWRWEZAGMEWKRYKMKRAYWRWRWCKWAGWAR 780
Qy 460 TATTATATATCTCTGCGCACTAAATAATTTCCCTTTTGGTGTATCTGGATGCT 519
Db 779 WMKSRYRWKXKYATRYWKNWMTWWSWRWKSYSRMGMRWMSAWRYCSRWCKAR 720
Qy 520 ACTCTGAAGTTAAGATAAAGACGATATAGTGAGACCTATGGAACCTTACACTATAAAT 579
Db 719 TKVASSARWTKAKRSYRYYRWYKRGNTYRYWRSCRWTRARYSK--RRKWAGAS 662
Qy 580 ATCACTATTATAATTAACGATAGGGAATATTCAGATATGTTTACTCAGTTCAAAAGT 639
Db 661 KSCWYWRGARSWYSKYSKSAKCKKTRYVTSYVSTGMYGYSYKMSWTSKMSYMG 602
Qy 640 GAGCTCGGTCGATCTTAACCTTCCTCCACTGGTGGGGGACATATATGGAAGAAAT 699
Db 601 KMTCTWYTNKGTRESKMGWSGMSMYMWWKMKREKMYMKWCKTWRRCMYRWG 542

CC immunodeficiency, trichiodystrophy, Fanconi's anaemia, solid tumours
CC and cancer, particularly by determining status of cytosine methylation
CC and/or by detecting single-nucleotide polymorphisms. Determination of
CC individual methylation patterns may allow development of individualised
CC therapies. The sequences given in records ABL92192-AB:92335 represent
CC chemically pre-treated DNA fragments from genes associated with DNA
CC repair, and their complements. Note: the sequence data for this patent is
CC not represented in the specification, but is based on sequence
CC information supplied by the European Patent Office
XX
SQ Sequence 10279 BP; 3242 A; 65 C; 1845 G; 5127 T; 0 U; 0 Other;

Query Match 4.5%; Score 49.4; DB 6; Length 10279;
Best Local Similarity 47.8%; Pred. No. 0.089;
Matches 143; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

QY 140 AACATAATATTAATAACCATATTACAGCATACAGTGAAGTCATCTCTGTATGATA 199
DB 7605 ATCATATAAATAAATTTAAATAAATAAACAATAAATAATTAATCCAAAAAACA 7546

QY 200 GGATGACTTTTTTATGTGTTCTTCACATACACTTAATGAGCATGTCACACAGTG 259
DB 7545 ATAAACCATTAATTAACCTATTTTTCACAAATATCCAAAAAATCTTACTTCTCAATA 7486

QY 250 AGATCCTAGCATGTTTCAGTCCAGCGGTGAAACAATATACATACAAATTTACGAAA 319
DB 7485 AAAAAACAATACTTAATTTCAACACGATAAAAAAATAAAAAATTAATAAAA 7426

QY 320 AAAGAAGTTTATAAAAAAGAGAGCTACAAATTAAGGCTATAAACAATATTTGTTCAAAA 379
DB 7425 ACATAAATACTAATATATATACAAATCCAACTTTTATCAACAAACAAATTAATAA 7366

QY 380 GTGTTAACTGCCATCCCGGCTTAACACTTAATCTAGCTCATTTTAACTGTATATAAAC 438
DB 7365 ACTATATCTTCATAAACACCCCAATATTTTATATATCTCATTTTAAATTCATTAAATAC 7307

RESULT 12
AAD22328/c
ID AAD22328 standard; DNA; 10279 BP.
XX
XX AAD22328;
XX
XX 12-FEB-2002 (first entry)
XX
XX Chemically treated human genomic DNA #18 associated with DNA adducts.
XX
XX DNA adduct; peptide nucleic acid; PNA; cytosine methylation;
XX Gene therapy; tumour; cancer; human; Gs.
XX
XX Homo sapiens.
XX
XX WO200177378-A2.
XX
XX 18-OCT-2001.
XX
XX 06-APR-2001; 2001WO-EP004015.
XX
XX 06-APR-2000; 2000DE-01019058.
XX 07-APR-2000; 2000DE-01019173.
XX 30-JUN-2000; 2000DE-01032529.
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-010923/01.
XX
XX Novel nucleic acid comprising sequence of a segment of chemically
XX pre-treated DNA of genes associated with DNA adduct, useful for diagnosis
XX and therapy of solid tumors and cancer.
XX

PS Claim 1; Page 63-68; 97pp; English.
XX
CC The invention relates to chemically modified DNA of genes associated with
CC DNA adducts, oligonucleotides and/or peptide nucleic acid (PNA) oligomers
CC for detecting cytosine methylations as well as method for ascertaining
CC genetic and/or epigenetic parameters of genes associated with DNA
CC adducts. Oligomers of the invention coupled to a solid phase is useful
CC for manufacturing an arrangement of different oligomers (array) fixed to
CC a carrier material for analysing diseases associated with the methylation
CC state of the CpG dinucleotides of chemically pretreated DNA of genes
CC associated with DNA adduct. They are also useful for ascertaining genetic
CC and/or epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases by analysing cytosine
CC methylations. Sequences of the invention are useful for diagnosis and
CC therapy of solid tumours and cancers. They are also useful in gene
CC therapy. The present sequence is chemically pretreated human genomic DNA
CC associated DNA adducts
XX
SQ Sequence 10279 BP; 3242 A; 65 C; 1845 G; 5127 T; 0 U; 0 Other;

Query Match 4.5%; Score 49.4; DB 6; Length 10279;
Best Local Similarity 47.8%; Pred. No. 0.089;
Matches 143; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

QY 140 AACATAATATTAATAACCATATTACAGCATACAGTGAAGTCATCTCTGTATGATA 199
DB 7605 ATCATATAAATAAATTTAAATAAATAAACAATAAATAATTAATCCAAAAAACA 7546

QY 200 GGATGACTTTTTTATGTGTTCTTCACATACACTTAATGAGCATGTCACACAGTG 259
DB 7545 ATAAACCATTAATTAACCTATTTTTCACAAATATCCAAAAAATCTTACTTCTCAATA 7486

QY 260 AGATCCTAGCATGTTTCAGTCCAGCGGTGAAACAATATACATACAAATTTACGAAA 319
DB 7485 AAAAAACAATACTTAATTTCAACACGATAAAAAAATAAAAAATTAATAAAA 7426

QY 320 AAAGAAGTTTATAAAAAAGAGAGCTACAAATTAAGGCTATAAACAATATTTGTTCAAAA 379
DB 7425 ACATAAATACTAATATATATACAAATCCAACTTTTATCAACAAACAAATTAATAA 7366

QY 380 GTGTTAACTGCCATCCCGGCTTAACACTTAATCTAGCTCATTTTAACTGTATATAAAC 438
DB 7365 ACTATATCTTCATAAACACCCCAATATTTTATATATCTCATTTTAAATTCATTAAATAC 7307

RESULT 13
ABK28407/c
ID ABK28407 standard; DNA; 9504 BP.
XX
XX ABK28407;
XX
XX 23-APR-2002 (first entry)
XX
XX DNA transcription associated genomic DNA #141.
XX
XX
XX DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
XX PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
XX single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
XX viral infection; Sezary syndrome; haematological disorder; tuberculosis;
XX immunological disorder; Werner syndrome; developmental disorder;
XX psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
XX neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
XX myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
XX angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
XX polyglutamine disorder; solid tumour.
XX
XX Unidentified.
XX
XX WO200192565-A2.
XX
XX 06-DEC-2001.
XX
XX 06-APR-2001; 2001WO-EP003973.
XX

XX 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
FA (EPIC-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-090046/12.
XX New nucleic acids or oligomers, useful for diagnosing or treating
PT diseases associated with DNA transcription, e.g. immunological disorders,
PT Werner syndrome, psoriasis, myocardial infarction, solid tumors or
PT cancer.
PS Claim 1; SEQ ID NO 281; 32pp; English.
XX The invention relates to a nucleic acid, which comprises a segment of the
CC chemically pretreated DNA of genes associated with DNA transcription from
CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC to the chemically pretreated DNA of genes associated with DNA
CC transcription. The set of oligomer probes are useful for detecting the
CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC diagnosing or treating diseases associated with DNA transcription
CC (particularly with the methylation status), e.g. adenosine deaminase
CC deficiency, viral infection, retroviral infection, Sezary syndrome,
CC haematological disorders, immunological disorders, Werner syndrome,
CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC neurological disorders, neurodegenerative disorders, Waardenburg
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
CC or cancer. Sequences ABK28121-ABK28472 represent DNA transcription
CC associated genomic DNA molecules of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification but
CC was obtained in electronic format directly from the European Patent
CC Office
XX Sequence 9504 BP; 2528 A; 138 C; 183 G; 4985 T; 0 U; 0 Other;
Query Match 4.4%; Score 48.2; DB 6; Length 9504;
Best Local Similarity 46.8%; Pred. No. 0.17;
Matches 152; Conservative 0; Mismatches 173; Indels 0; Gaps 0;
QY 317 AAAAAGAGCTTTAATAAAGAGAGAGCTACAATTAAGCGCTATAACAATTTGTTCA 376
DB 9072 AACATAAAATTTCTAAATAATTAATAAATAATTTAACTAAATATAAATAAATAAATAA 9013
QY 377 AAAGTGTAACTGCCATCCGCTCAACCTTAACCTCACTCACTCACTCACTCACTCACTCACT 436
DB 9012 AANTAGACACAAAACAACTTAACAATTAATTAATAAATAAATAAATAAATAAATAAATAA 8953
QY 437 ACGCGGCTTCAGGTGCAAGTTTATATATATATATATATATATATATATATATATATATAT 496
DB 8952 AAACAAAACGTTTAACTAAATTTATATAAATAAATAAATAAATAAATAAATAAATAAATA 8993
QY 497 CTTTTGTGTAICTGGATGCTACTCTGAGTGAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 556
DB 8892 TTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 8833
QY 557 CCTATGGAACCTACATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 616
DB 8832 AATCACTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 8773
QY 617 TATGGTTACCTCAGTTCAAAAGTGA 641
DB 8772 AATATATATCTTCATTTCTAATAATA 8748

RESULT 14
AAA70139
ID AAA70139 standard; DNA; 5139 BP.
AC AAA70139;
XX 07-NOV-2000 (first entry)
DT Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:272.
DE Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX antimalarial; malaria; protozoacide; infection; insecticide; db.
XX Plasmodium falciparum.
OS WO2000025728-A2.
PN 11-MAY-2000.
PD 05-NOV-1999; 99WO-US026796.
XX 05-NOV-1998; 98US-0107131P.
PR (HOFF/) HOFFMAN S.
XX (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX Hoffman S, Carucci D, Gardner M, Venter JC;
XX WPI; 2000-365347/31.
XX Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection.
XX Disclosure; Page 482-484; 577pp; English.
PS The present invention describes proteins and their fragments (I) encoded
XX by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II). (I) and
CC (II) are useful for the development of vaccines against P. falciparum
CC infection. (I) and polyclonal antisera or a monoclonal antibody raised to
CC immunogens comprising the sequences of (I), are useful in the detection
CC of infection with P. falciparum. Furthermore, (I) (especially when they
CC are rifins or secreted or membrane proteins) can aid the identification
CC of drugs to treat or prevent P. falciparum infection, or they can be used
CC to identify drug resistance in P. falciparum. Sequencing of the
CC Plasmodium chromosome 2 and the subsequent identification of proteins
CC encoded by it will help to expand our understanding of parasite biology,
CC a process hampered by the complexity of the parasitic lifecycle, and
CC provide new targets for vaccine and drug development. Parasite resistance
CC to drugs and mosquito resistance to insecticides have led to a resurgence
CC of malaria in many parts of the world, and there is a pressing need for
CC vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352
CC represent nucleotide and protein sequences given in the present
CC invention, but which are not specifically mentioned within the
CC specification
XX Sequence 5139 BP; 2440 A; 434 C; 712 G; 1553 T; 0 U; 0 Other;
Query Match 4.3%; Score 47.2; DB 3; Length 5139;
Best Local Similarity 44.5%; Pred. No. 0.25;
Matches 187; Conservative 0; Mismatches 233; Indels 0; Gaps 0;
QY 193 TATGATAGATGATCTTTTATGTTGTTCTTCAACATACACTTAATGAGCATGTCCA 252
DB 997 TATTATAATAATATATTTTAAATCCATTAATGATAATTAATCAATTAATGTTGTTATACA 1056
QY 253 ACCAGTGCAGAACCTAGCAGTTCATCGGTGCGGTCAGCAACAATATACATTACATTT 312
DB 1057 ACTTGTGATATAAAGAAAGTTATCGAAACAGCAAGTGATATGATAACAAATTTTT 1116

313	ACG	AAAAAAGAGAGCTTTT	ATAA	AAAGAGAGCGCT	CAAA	TTAA	AGGCTATA	AAACAA	TATT	ATTTG	372
1117	GAA	AGATGA	TGAAT	GAGAAAAA	AAAT	TATAT	AAAAAT	GAAAAAT	TTTAA	TGATGTTAT	1176
373	TT	CAAAAGTGT	TAACT	CGGCGCT	TAAC	TAACT	CAGCT	CATT	TTAACT	GTGAAT	432
1177	AAAA	AAAAAAT	TCA	NTGG	CAAA	GAAAAA	TACT	TAA	TTCAA	TAGTACA	1236
433	AAAA	ACGGCGCT	TCAG	GTGCA	GT	TTAT	TTAT	TTAT	TC	TGTCGCACT	492
1237	AAA	AGGTTTC	TTTT	TAA	TTCT	AGGA	TGAG	CACT	TAT	TTATG	1296
493	TT	CGCTTT	TGGT	GTACT	CTCG	AGT	TAAG	GTAA	AAAG	CAAGATAT	552
1297	TATA	ATAT	TGTT	TAAG	AGGG	AAAA	CC	GGA	AGAT	ATAA	1356
553	GAG	ACCTAT	TGGA	CTTAC	ATA	TATAT	CAC	TATT	AAAT	TAACTG	612
1357	AAA	AGCAAT	TAAT	TAAT	TGTA	TAAT	TAAT	TAAT	TAAT	TAAT	1416

RESULT 15

ABLI0856/C

ID ABL10866 standard; cDNA: 4176 BP.

AA ABL10866;

AA	DT	26-MAR-2002	(first entry)

XX DE *Drosophila melanogaster* expressed polynucleotide SEQ ID NO 27080.

KX Drosophila; developmental biology; cell signalling; insecticide;
KW Drosophila; developmental biology; cell signalling; insecticide;

KW **pharmaceutical; gene; ss.**
KW **pharmaceutical; gene; ss.**

AA Drosophila melanogaster.

XX PN WO200171042-A2.

XX
PD
27-SEP-2001.XX
PF
23-MAR-2001: 2001WC-IIS009231

XX
PP 23-MAR-2000. 2000TS-0191637D

PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.

XX
PA (PEKE) PE CORP NY.

XX PI Venter JC. Adams M. Li PWD. Myers EW.

XX
DR
WPI: 2001-656860/75

DR WFI; ZOUT=656860/
DR P-PSDB: ABB66763.

xx New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from *Drosophila* and for elucidating cell signaling and cell-cell
PT interactions.

PS Claim 1; SEQ ID NO 27080; 21bp + Sequence Listing: English.

The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from *Drosophila*. The invention is useful in developmental biology and in elucidating cell signaling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB116176-AB130511), expressed DNA sequences (AB101840-AB116175) and the encoded proteins (AB057737-AB572072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from Wipo at ftp.wipo.int/pub/published pct sequences

SQ Sequence 4176 BP; 1186 A; 862 C; 837 G; 1291 T; 0 U; 0 Other;

Query Match 4.28; Score 45.2; D3 4; Length 4176;

Best Local Similarity	56.8%;	Pred. No. 0.73;			
Matches 103;	Conservative	0;	Mismatches	78;	Indels
					1; Gaps
					1;
QY	689	TTGGAGAAANTCTGTGATCTGCTTTTATGATGGATATAGTATTAACAGCAGCTCTT	748		
DB	2504	TCGAAAGCAAATATTTTAAATATGCTTCGGTATGGGTAAACATCAAAAAGAGCANITTT	2445		
QY	749	TGCAGATAGATTTCAGGATAACAACTCCATAAATCTGATCGGAAATTTTACTTAGGAAA	808		
DB	2444	AGCAATACGCATCGATGCTAATAACCCGTGCTCCGTTCCGAATATATCAATAATGAC	2385		
QY	809	TAAATGATGACACCAAGAAATTCGTATACTTTGTCACTTCTTCGGCGGGTAAAGT	867		
DB	2384	TAAATTAAGAACTAAGAACTTTAAATATCCTCTAATTAATCTCTTCGCGAGAAAAAT	2325		
QY	868	TT	869		
DB	2324	TT	2323		

Search completed: May 8, 2004, 09:40:48
Job time : 495 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2004, 09:32:30 ; Search time 3220 Seconds
(without alignments)
10071.530 Million cell updates/sec

Title: US-09-839-894-9

Perfect score: 1086

Sequence: - atgaataagattttttttt.....caagtagtcaaacactcttag 1086

Scoring table: IDENTITY NJC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:**

- 1: em_estba:**
- 2: em_esthum:**
- 3: em_estlin:**
- 4: em_estmu:**
- 5: em_estov:**
- 6: em_estpl:**
- 7: em_estro:**
- 8: em_htc:**
- 9: gb_est1:**
- 10: gb_est2:**
- 11: gb_htc:**
- 12: gb_est3:**
- 13: gb_est4:**
- 14: gb_est5:**
- 15: em_estfun:**
- 16: em_estom:**
- 17: em_gss_hum:**
- 18: em_gss_inv:**
- 19: em_gss_pln:**
- 20: em_gss_vrl:**
- 21: em_gss_fun:**
- 22: em_gss_mam:**
- 23: em_gss_mus:**
- 24: em_gss_pro:**
- 25: em_gss_fod:**
- 26: em_gss_png:**
- 27: em_gss_vrl:**
- 28: gb_gss1:**
- 29: gb_gss2:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	59.4	5.5	1101	CNS0100X	AL098379 Drosophila
	55	5.1	1056	13	EX415058
	52	4.8	1200	13	EX415878
c 4	51.6	4.8	1101	29	CNS0039G
					AL063921 Drosophila

5	51.2	4.7	1067	14	CD386564	AGENCOURT
6	51	4.7	1201	29	CNS00RVE	AL071298 Drosophila
7	51	4.7	1225	13	CNS0161D	AL106171 Drosophila
8	50.8	4.7	1124	13	EX346282	EX436282
9	50	4.6	1179	13	EX380916	EX380916
10	49	4.5	1043	29	CNS0145P	AL103735 Drosophila
11	49	4.5	1061	13	EX437039	EX437039
12	48.6	4.5	1201	13	EX446296	EX446296
13	48.4	4.5	1201	13	EX356554	EX356554
14	48	4.4	1101	29	CNS0039G	AL063921 Drosophila
15	47.6	4.4	931	14	CD389151	CD389151 AGENCOURT
16	47.6	4.4	999	13	EX380865	EX380865
17	47.6	4.4	1101	29	CNS0145U	AL103740 Drosophila
18	47.6	4.4	1148	13	EX360752	EX360752
19	47.4	4.4	925	28	BH152177	BH152177 ENTOD23TF
20	47.4	4.4	1101	29	CNS000D1	AL065414 Drosophila
21	47.2	4.3	1101	29	CNS014XJ	AL104737 Drosophila
22	47.2	4.3	1204	29	CNS016E2	AL106628 Drosophila
23	47	4.3	838	29	CNS0096O	AL053091 Drosophila
24	47	4.3	1101	29	CNS00PMC	AL070972 Drosophila
25	47	4.3	1194	13	EX462185	EX462185
26	46.8	4.3	536	13	BQ400483	EQ400483 NISC.mp09
27	46.8	4.3	1101	29	CNS016HF	AL106743 Drosophila
28	46.6	4.3	408	10	BE224677	BE224677 kp59h11.Y
29	46.6	4.3	460	12	BI772835	BI772835 kp44h09.Y
30	46.6	4.3	1201	9	AL532454	AL532454
31	46.6	4.3	1201	13	EX343343	EX343343
32	46.4	4.3	812	28	AZ529785	AZ529785 ENTDA08TF
33	46.4	4.3	911	28	AZ687106	AZ687106 ENT171TR
34	46.4	4.3	1092	29	CNS020K7	AL175696 Te-raodor
35	46.4	4.3	1101	29	CNS000B8	AL063632 Drosophila
36	46.4	4.3	1101	29	CNS00EVL	AL069706 Drosophila
37	46	4.2	427	13	EX403499	EX403499
38	46	4.2	1026	13	EX415239	EX415239
39	45.8	4.2	1201	9	AL536104	AL536104
40	45.6	4.2	852	29	CNS009AX	AL053547 Drosophila
41	45.6	4.2	928	29	CNS00DKY	AL071865 Drosophila
42	45.2	4.2	903	29	CNS0173V	AL108349 Drosophila
43	45.2	4.2	994	13	EX414650	EX414650
44	45.2	4.2	1080	29	CNS005PF	AL069494 Drosophila
45	45	4.1	764	14	CF289381	CF289381 AGENCOURT

ALIGNMENTS

RESULT 1
CNS0100X/c 1101 bp DNA linear GSS 26-JUL-1999
LOCUS
DEFINITION Drosophila melanogaster genome survey sequence Sp6 end of BAC
BACN03G04 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL098379
VERSION AL098379.1 GI:5603990
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : secre@genoscope.cns.fr - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaut at C3PH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payar. It has been constructed in the vector

AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
Bp 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0CAP009BB02QPI.

FEATURES
source Location/Qualifiers
1..1200
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/cloned="CS0CAP008YIC4"
/tissue_type="THYMUS"
/clone_lib="Homo sapiens THYMUS"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(GT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."

ORIGIN

Query Match	4.8%; Score 52; DB 13; Length 1200;
Best Local Similarity	35.9%; Pred. No. 6.1;
Matches	229; Conservative 75; Mismatches 331; Indels 3; Gaps 2;

```

Qy      4  AATPAAGATTGTTATTATTTACATTCGTCTTCCTCCAGTACTTTTACATTTGCCTGTA 63
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     401  AADCTTATWTTWECCCATSTTSCITTTMAAAATTTTBBSWTATATTTATTTTATTA 460

Qy      64  TCGCAGATAAATTCGGGAGATGAAGCAATACTAATATTTTTGGCCGCCGTGACAGG 123
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     461  TAAATTTATGWATTTAATATATTTTTTAAAATTTTAAATTTTAAAAATTTWATAWWTT 520

Qy     124  AACGAATCTCCCCCAAACAATAATATTAATAAACONTATACAGATACAGTGAAAGT 183
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     521  TWAWATTTTTTAATWAAATTAATWAATTTTTTTTTTAATAATTTTAAAAAATAATWATTT 580

Qy     184  CATACTCTGTATGATAGATGACTTTTATCTTGTGCTCTCACATACACTTAATGGA 243
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     581  TWTTTTTWTAANAAMWTTTATATTTTWTWTTTATTTTATTTTATTAATWTTTATAAA 640

Qy     244  GCATCTCCAACCAGTGA--GAATCCTAGAGTTCATCGGTGAGCGGTGAARCAATATAA 301
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     641  ATWAATAATAATTAATTAATWTRAWAANTWATAATWTTTTTAAATAGATTAAATATAAT 700

Qy     302  CATTACAATTTACGGAAAAAAGAAGTTTAATAAAAGAGAGCTACAATAAAGCTATA 361
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     701  RAATTTATTTWAATTTWAATWWWAAWTMTTWAAGAAATTAATACATANATATATA 760

Qy     362  ACAATATTGTTCAAAAGTGTTAATCGCCATCCGGCTCAACTCACTCACTCAGTCAAT 421
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     761  TAWDCATRAAATATWATWTAATATAATTTWAAATWACAAANATWAAAAAATTTAA 820

Qy     422  TTAACTGTAAATAAAACGGCGTCTCAGGTGCAAGTTTATATTTATATATTCCTGTCGG 481
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     821  AAAATTTWAAATWAAAAAANAATTTTAAAAAANAATTTATTTTAAATTTWATWAAATTA 880

Qy     482  AACTAAAAAATTTGCCCTTTTGGTGGTATCTCGGATGCTACTCTGAAGTTAAGAGTA 541
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     881  TAAATPAAATATASAAAAATTTTTTTWTTATTTATTTTAAATTTWAAATTTTTTTTNAWA 940

Qy     542  GACGATATAGTGAGACCTTGGNACTTACACTATAAATATCATATATABATTAACCTGTA 601
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     941  -ATFAAATAWNAWAAAAAATAATTTWAAAWTATAAATAANTANWTAAWATWAAAAATWW 999

Qy     602  AGGGAATAATTCAGATATGGTTACCTCAGTTTCAAAAAGT 639
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db    1000  ATWAAWNTTTAAAAATWTATATTTTWTWTTAAAAAATAAATTT 1037

```


[illegible]

```

RESULT 7
CNS0161D      1225 bp      DNA      linear      GSS 26-JUL-1999
LOCUS         Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION    BACN15C18 of DrosBAC library from Drosophila melanogaster (fruit
               fly) , genomic survey sequence.
ACCESSION     AU106171
VERSION       AU106171.1  GI:5620504
KEYWORDS      GSS.
SOURCE        Drosophila melanogaster (fruit fly)
ORGANISM      Drosophila melanogaster
               Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
               Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
               Ephydroidea; Drosophilidae; Drosophila.
REFERENCE     1. (bases 1 to 1225)
               Genoscope.
AUTHORS       Direct Submission
TITLE         Submitted (23-JUN-1999) Genoscope - Centre National de Sequences :
               BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
               - Web : www.genoscope.crs.fr)
COMMENT       Determination of this BAC-end sequence was carried out as part of a
               collaboration with the European Drosophila Genome Project (EDGP) -
               http://www.edgp.sbi.ac.uk -. This Drosophila melanogaster BAC
               library (Dros BAC) was made by Alain Billaut at CEPH (Centre
               d'Etude du Polymorphisme Humain) with funding provided by a MRC
               project grant. The DNA was prepared from embryos by Alain Bucheton
               and Genevieve Payan. It has been constructed in the vector
               pBeloBAC11.
FEATURES             Location/Qualifiers
             1..1225
                /organism="Drosophila melanogaster"
                /mol_type="genomic DNA"
                /db_xref="taxon:7227"
                /clone="BACN15C18"
                /clone_lib="DrosBAC"
                /plasmid="pBeloBAC11"
                /note="end : SP6"
FEATURES             source

```

```

ORIGIN
Query Match.          4.7%; Score 51; DB 29; Length 1225;
Best Local Similarity 34.9%; Pred. No. 9.1;
Matches 119; Conservative 64; Mismatches 154; Indels 4; Gaps 1;

      288 TGAACCAATATAAACAATTACAATTTACGGAAAAAAAGAACGTTTAATAAAAAGACAGCTACA 347
        ||| ||| ||| ||| ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
      881 TAATAAAAAAAAAAAAAAWWATITTYTYYAYAWAAAYAWATAAANWATWTTTAAAWAAASA 940
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      348 AATTAAGAAGGTATAAACAANTATTGTCAPAAAGTGTTAACTGCCCATCCCGCGCTAACACT 407
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      941 AAAAANAANWTTTVAARAAAANWTWAAATAAATATTTTTTTTAAWWMTTATATWATAATT 1000
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      408 TAAC TGCGTCATTTTAAACUGTAATAAATAAACGGCGGCTTCAGGTGCAGTTTATATCTTATA 467
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      1001 TATWTTTWTWTWTWTWTAAATAAATAAATAATATTTTWTWANWMTWTWATTTTTTATA 1060
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      463 TAITCCTGCTGGCGAAGCTAAABAAATTTGCGTTTGGTGATATCTGGGA-----TGCTACTC 523
        ||| ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
      1061 TATWATATTTTATATTWANAWAWATWTWTWTWTWTWTWTWTWTWTWTWTWTATATTTTWT 1120
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      524 TGAAGCTTAAAGATAAAAGACGATATAGTGAGACCTATAGGAACCTTACACTATATAAAATCA 583
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      1121 TATWTAANATWATATAAABAATTTTWWATAATAATATWATATATATTTTATATAAANAANW 1180
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      584 CTATTAATAATTAACGTGAAGGGAAATATTCAGATAATGGTTA 624
        ||| ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
      1181 TTCTWTWTWTWTWTWTWAAATAAATAAATAAATAATWATWA 1221
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
BX436282/c linear EST 15-MAY-2003
LOCUS BX436282 Homo sapiens THYMUS Homo sapiens cDNA clone CSOCAF001.YC01
DEFINITION 5'-PRIME. mRNA sequence.

```

```

SOURCE      Homo sapiens (human)
ORGANISM:   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1124)
AUTHORS     Li W.B., Gruber C., Jessee J. and Polayes D.
TITLE       Full-length cDNA libraries and normalization
COMMENT      Unpublished (2001)
CONTACT     Genoscope
GENOSCOPE   Genoscope - Centre National de Séquençage
            BP 191 91006 EVRY cedex - France
            Email: secre@genoscope.cns.fr
            Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ Invitrogen corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0CAP001AH010P1.
FEATURES
            Location/Qualifiers
            1..1124
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /clone="CS0CAP001YCC1"
               /tissue_type="THYMUS"
               /clone_lib="Homo sapiens THYMUS"
               /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
               with a NotI-cligo(dT) primer. Five prime end enriched,
               double-strand cDNA was digested with Not I and cloned into
               the Not I and EcoRV sites of the pCMVSPORT 6 vector.
               library was not normalized."

```

```

ORIGIN
      Query Match      4.7%; Score 50.8; DB 13; Length 1124;
      Best Local Similarity 21.3%; Pred. NO. 10;
      Matches 138; Conservative 197; Mismatches 313; Indels 0; Gaps 0

```


BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 7009.f For
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CSOCAP004D02QP&cluster=7009.f. Contact :
 Peng Liang Email: fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID : CSOCAP004D02QP1.
 Location/Qualifiers
 1..1061
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CSOCAP004YH04"
 /tissue_type="THYMUS"
 /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
 with a NotI-oligo(dT) primer. Five prime end enriched,
 double-strand cDNA was digested with Not I and cloned into
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.
 Library was not normalized."

BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 7009.f For
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CSOCAP004D02QP&cluster=7009.f. Contact :
 Peng Liang Email: fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID : CSOCAP004D02QP1.
 Location/Qualifiers
 1..1061
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CSOCAP004YH04"
 /tissue_type="THYMUS"
 /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
 with a NotI-oligo(dT) primer. Five prime end enriched,
 double-strand cDNA was digested with Not I and cloned into
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.
 Library was not normalized."

Query Match 4.5%; Score 49; DB 13; Length 1061;
 Best Local Similarity 42.2%; Pred. No. 22; Indels 0; Gaps 0;
 Matches 139; Conservative 17; Mismatches 173;
 QY 288 TGAAACAAATATAACATTACAAATTTACGGGAAAAAGAGTTTATATAAAGAGAGCTACA 347
 DB 515 TAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 574
 QY 348 AATTAAGGCTATAACATATTATTCTCAAGCTGTTAACTGCCCTCCGGCCCTAACACT 407
 DB 575 WAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 634
 QY 408 TAACCTAGCTCTATTTAACTGTAATAAAGCGGCTTCAGGTGCAAGTTTATATTATATA 467
 DB 635 TATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 694
 QY 468 TATTCCTGCTGGGAACTAAAAATTTTGGCTTTTGGTGTATCTGGGATGCTCTCTGAA 527
 DB 695 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 754
 QY 528 GTTAAGAGTAAAGAGGATATAGTGGACCTATCGNACTTACACTTAATATCACTAT 587
 DB 755 WTAT 814
 QY 568 TAAATTAACGTAAAGGAAATATTTCAGA 616
 DB 815 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 843

Query Match 4.5%; Score 49; DB 29; Length 1043;
 Best Local Similarity 37.9%; Pred. No. 23; Indels 3; Gaps 1;
 Matches 136; Conservative 48; Mismatches 172;
 QY 11 TTTTATTATTATTACATTTTCTTCCTCAGTACTTTTATTTTATTTTATTTTATTTTATTT 470
 DB 881 TTTTATTATTATTACATTTTCTTCCTCAGTACTTTTATTTTATTTTATTTTATTTTATTT 822
 QY 71 ATAAATTCGGAGAGCAAGCAATCAATCAATTTTGGCCCGCGTGACAGCAACCAAT 130
 DB 821 AAAAAATTTTWTWW 762
 QY 131 CTTCCTCCCAACATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 190
 DB 761 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 702
 QY 191 TGTATGATAGGATGACTTTTATTTTATTTTGTCT--TCTCAATACACTTATTTGAGCAT 247
 DB 701 AATAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 642
 QY 248 GTCAACACGTGAGATCTTACGAGTTCAGTCCGTCAGCGGTGAACAAATATACATTAC 307
 DB 641 ATAAAWAWWWWWAAAAAATAATATATTTTATATATATATATATATATATATATATAT 582
 QY 308 AATTTCGGAAAAAGAGTTTATATAAAGAGAGCTACAAATTTAAAGGCTATATAACAA 366
 DB 581 TTTAT 523

RESULT 12
 BX446296 1201 bp mRNA linear EST 15-MAY-2003
 LOCUS BX446296 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
 DEFINITION XLOC000012A03 3-PRIME, mRNA sequence.
 ACCESSION BX446296
 VERSION BX446296.1 GI:30784407
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1201)
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France

RESULT 11
 BX437039 1061 bp mRNA linear EST 15-MAY-2003
 LOCUS BX437039 Homo sapiens THYMUS Homo sapiens cDNA clone CSOCAP004YH04
 DEFINITION 5-PRIME, mRNA sequence.
 ACCESSION BX437039
 VERSION BX437039.1 GI:30789546
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1061)
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2004, 08:42:24 ; Search time 98 Seconds
(without alignments)
6149.762 Million cell updates/sec

Title: US-09-839-894-9
Perfect score: 1086
Sequence: 1 atgaataagattttattat.....caagtagtcaaacactctag 1086

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/prodata/2/ina/5A COMB.seq:
2: /cgn2_6/prodata/2/ina/5B COMB.seq:
3: /cgn2_6/prodata/2/ina/6A COMB.seq:
4: /cgn2_6/prodata/2/ina/6B COMB.seq:
5: /cgn2_6/prodata/2/ina/PCITUS COMB.seq:
6: /cgn2_6/prodata/2/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	298.8	27.5	5798	2	US-08-483-101-1
2	43	4.0	832	4	US-09-621-976-2813
3	39.8	3.7	74962	4	US-09-685-853A-3
4	39	3.6	10640	4	US-09-417-4850-5
5	38.8	3.6	342	4	US-09-601-198-115
6	38.4	3.5	4171	4	US-08-956-1715-469
7	38.2	3.5	473	1	US-08-764-100-16
8	38.2	3.5	4970	1	US-08-764-100-14
9	38.2	3.5	4970	1	US-08-764-100-20
10	38.2	3.5	1664976	4	US-08-916-421B-1
11	37.8	3.5	832	4	US-09-621-976-2813
12	37.6	3.5	5562	4	US-10-204-708-63
13	37.4	3.4	533	4	US-09-495-050A-48
14	37.2	3.4	7075	4	US-08-956-1715-263
15	36.6	3.4	2671	6	5168051-9
16	36.4	3.4	1020	4	US-09-463-962-1
17	36	3.3	540	4	US-09-404-879A-2
18	36	3.3	540	4	US-09-338-933-2
19	36	3.3	540	4	US-09-215-681-2
20	36	3.3	540	4	US-09-216-003A-2
21	36	3.3	551	4	US-09-404-879A-137
22	36	3.3	551	4	US-09-338-933-137
23	36	3.3	551	4	US-09-215-681-137
24	36	3.3	551	4	US-09-216-003A-137
25	36	3.3	15016	4	US-09-601-198-60
26	36	3.3	319608	4	US-09-539-333D-1
27	36	3.3	319608	4	US-09-679-403-1

C 28	35.6	3.3	201	4	US-08-956-171E-2594
C 29	35.6	3.3	1689	1	US-07-991-867B-41
C 30	35.6	3.3	1689	2	US-08-544-332-41
C 31	35.6	3.3	1689	4	US-09-370-861A-41
C 32	35.6	3.3	1947	4	US-09-370-861A-74
C 33	35.6	3.3	8457	1	US-07-991-867B-1
C 34	35.6	3.3	8457	2	US-08-544-332-1
C 35	35.6	3.3	8457	4	US-09-370-861A-1
C 36	35.6	3.3	19124	2	US-08-487-826B-13
C 37	35.4	3.3	2826	4	US-08-624-655A-1
C 38	35.4	3.3	2985	4	US-09-833-381-1138
C 39	35.4	3.3	162450	4	US-09-345-882-1
C 40	35.4	3.3	169998	4	US-09-676-610B-24
C 41	35.4	3.3	197496	4	US-09-877-177A-1C
C 42	35	3.2	1491	4	US-09-134-001C-866
C 43	35	3.2	11770	4	US-08-961-527-172
C 44	35	3.2	640681	4	US-09-790-988-1
C 45	34.8	3.2	337	2	US-09-032-684-11

ALIGNMENTS

RESULT 1
US-08-483-101-1
; Sequence 1, Application US/08483101
; Patent No. 5932715
; GENERAL INFORMATION:
; APPLICANT: Scott, June R.
; APPLICANT: Froehlich, Barbara
; APPLICANT: Caron, Judy
; TITLE OF INVENTION: CS2 Proteins and Coding Sequences
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee and Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,101
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33875
; REFERENCE/DOCKET NUMBER: 6-95
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5798 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: Gouble
; TOPOLOGY: Circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; FEATURES:
; NAME/KEY: CDS
; LOCATION: 499..1215
; FEATURE:
; NAME/KEY: sig peptide
; LOCATION: 499..552


```
; LENGTH: 342
; TYPE: DNA
; ORGANISM: Ureaplasma urealyticum
US-09-601-198-115

Query Match          3.6%; Score 38.8; DB 4; Length 342;
Best Local Similarity 51.8%; Pred. No. 0.33; Mismatches 82; Indels 0; Gaps 0;
Matches 88; Conservative

QY 684 ATATTGGAAGAAATTCGTTGATATGTCGTTTATGATGATATAGTACTTAACAGCAG 743
DB 303 ATAAATTCGAATATAAATTCGATATATGCTTCTTAATGATTTAGAAATCCCAATTAA 244
QY 744 CTCCTTGGAGATAGATTCAGGATACAAATCCTAAATCTGATGGAATTTATCTTAAG 803
DB 243 ATCATCTTAGTAAAGTTAAATTTAAGATAGCTAAATTTGATTTTAAAAAT 184
QY 804 GAAAAATAATGATGACACCAAGAAATTCATATATCTTTGTCACCTTCTCT 853
DB 183 GAATAAAGTGGTCTATATATTTTACTTTCGTAAATGTTTACCTTTT 134

RESULT 6
US-08-956-171E-469
; Sequence 469, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171E
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB24071
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 469:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4171 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 469:
US-08-956-171E-469

Query Match          3.5%; Score 38.4; DB 4; Length 4171;
```

```
Best Local Similarity 54.2%; Pred. No. 1.1;
Matches 78; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 803 GGAAATAATGATGACACCAAGAAATTCGATATATCTTTCACCTTCTCTTGGCGGTA 862
DB 3932 GTACAAATAGAAACACCAATACACAAACCCCAAAATATGATATATTTTTCAGTAA 3991
QY 863 AAGTTTAACTCCACAAATGGAAGTCATTAATTAATTTGCTGACGCAGCTTCTCTGGAAA 922
DB 3992 ACAGTCGAATAAGACCAAAAGTAACGGATATTAATCCTACAAAATAGTTGCCATCACA 4051
QY 923 CAAGCTGGAATAGATTACAGCTG 946
DB 4052 TAAATAGATAGCGCTATATATG 4075

RESULT 7
US-08-764-100-16
; Sequence 16, Application US/08764100
; Patent No. 5773700
; GENERAL INFORMATION:
; APPLICANT: van Grinsven J., Martinus Q.
; APPLICANT: De Haan, Petrus T.
; APPLICANT: Gieles L., Johannes J.
; APPLICANT: Peters, Dirk
; APPLICANT: Goldbach, Robert W.
; TITLE OF INVENTION: Improvements in or Relating to Organic
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sandoz Agro, Inc
; STREET: 975 California Avenue
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,100
; FILING DATE: 06-DEC-1996
; CLASSIFICATION: 80C
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/214,064
; FILING DATE:
; APPLICATION NUMBER: US 08/032,235
; FILING DATE: 17-MAR-1993
; APPLICATION NUMBER: GB 9206016.9
; FILING DATE: 19-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: No 5773700ris, Allen E.
; REGISTRATION NUMBER: 34,450
; REFERENCE/DOCKET NUMBER: 137-1061
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 354-3592
; TELEFAX: (415) 857-1125
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 473 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-764-100-16

Query Match          3.5%; Score 38.2; DB 1; Length 473;
Best Local Similarity 43.8%; Pred. No. 0.53;
Matches 166; Conservative 0; Mismatches 213; Indels 0; Gaps 0;

QY 420 TTTTAACTCTAATAAACAAGCGGCTTCAGTGCAGTTTATATATATATATATATCTCTCTGG 479
```



```
; PILING DATE: 19-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5773700ris, Allen E.
; REGISTRATION NUMBER: 34,490
; REFERENCE/DOCKET NUMBER: 137-1061
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 354-3592
; TELEFAX: (415) 857-1125
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4970 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-764-100-20

Query Match          3.5%; Score 38.2; DB 1; Length 4970;
Best Local Similarity 43.8%; Pred No. 1.3;
Matches 166; Conservative 0; Mismatches 213; Indels 0; Gaps 0;

Qy 420 TTTTAACTGTATATAAAACGGCGGTCAGGTGCAAGTTTATATTATATATATTCCTGCTGG 479
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 480 CGAAGTAAATTTCCCTTTTGGTGGTATCTGGGATCTCTCTGAAGTTAAGAGTTAA 539
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3867 AGTATAAATATATATGATGATTTGTTTAAACCAAAATCAAAAACCAAAAAGAAAAAG 3808
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 540 AAGACGATATAGTGAGACTTGGAGCTTACACTATAATATCACTATTAATTAATTAACGTA 599
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3807 AAAAAATAAACAAAAACAAAAACAAAAACAAAAACAAAAACAAAAACAAAAATAGAAAA 3748
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 600 TAAGGAATATTCAGATATATGTTACCTCAGTTCAGTTCAAAAGTCAGCTCGATCTTAA 659
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3747 AAGTTGAAAAAACCACAAAAATTTTGTAAATAAATAGGCTCGGCCAGATTGG 3698
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 660 CTTCGCTCAACTGTGGGGCAGCATATATGGAGAAATCTGTGTATATGCTTTTA 719
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3687 TCTAAGACCTTTTATTTGTTTATACATTTTATTTGTTGTTGTTTATTTTATTTTA 3628
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 720 TGATGCATATAGTACTAACAGCAGCTCTTTCGAGATAAGATTTCAGGATAACAATCTTAA 779
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3627 TTATTTTATATTTTATATATAGTTTGCTTATTTAACACTTATTTAGACAAATTAATTT 3568
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 780 ATCGATCGGAATTTTAT 798
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3567 ATTGATTACAATCAATCT 3549
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Jannaschli
; Patent No. 6503729
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/524,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschli
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
```

```
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
```

```
; LOCATION: (559241)..(559241)
; NAME/KEY: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1084930)..(1084930)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
```

```
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
; US-08-916-421B-1

Query Match      3.5%; Score 38.2; DB 4; Length 1664976;
Best Local Similarity 51.5%; Pred. No. 11;
Matches 88; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

Qy 747 TTTCGAGATAGATTTCAGGATACACATCCATAATCTGATGGGAAATTTATCTAAGAA 806
Db 1486909 TGTGAATTTGAATTAAGAAACAAATCCTGTTTATGTTGAAAAATCCTATGAAAACT 1486968
Qy 807 AATAAATGATGACACCAAGAAATTCATATATCTTGTCACTTCTTGGCGGTAAAAG 866
Db 1486969 CTTAGCTGATATACAAAGATGATGCAATCTATTGAACCTAGGATTTGAGGGCGCTT 1487028
Qy 867 TTAACTCCAAAGAAATGGAACGTCATTAATAATTTGTCGACGACCTTCTCT 917
Db 1487029 TTTAAATAAAACAGTATATCTTTATTATGAAAAATGACGAAATCATCT 1487073

RESULT 11
US-09-621-976-2813/c
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2813
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 235..399
; US-09-621-976-2813

Query Match      3.5%; Score 37.8; DB 4; Length 832;
Best Local Similarity 13.4%; Pred. No. 0.85;
Matches 49; Conservative 162; Mismatches 159; Indels 3; Gaps 1;

Qy 568 TACACTATAATATCACTATTAAATTAAGTGAAGGGAATATTCAGATATGTTACCT 627
Db 370 TATTTTGTYYWKKTWYTWYTWYTWYTWYTWYTWYTWYTWYTWYTWYTWYTWYTWY 311
Qy 628 CAGTTCAAAAGTGACGCTCGGTCGATCTTAACCTTCGTCCTCAACTCGTGGGGGCAATAT 687
Db 310 WYWKRMWSTWYCYWCKCCMYRRCACAWTWARGRWMSYAWGKWSMESAYSMCTRY 251
Qy 688 ATTGGAAGAAATCTCTTGTATATGCTTTTATGATGATATAGTACTACACAGACCTCT 747
Db 250 YRKGSTYWTMKCTCATWYWKYKRWMSKTCWSGSGGYMTSTSTSYWYASWY 191
Qy 748 TTGGAGATAAGATTTTCAGGATAACATCCTAAATCTGATGGGAAATTTTATCTTAAGGAA 807
Db 190 TMCWWRWSTWYWYWKWKKWYATTTWRAMWYWA---ATWYWYWYWYWYWYWYWY 134
Qy 808 ATAATGATGACACCAAGAAATTCATATATCTTTGTCACTTCTTGTGGGGTAAAGT 867
Db 133 AMYRRIMMWGYWYWKYSYRTRCAWAYAKTKGYSYWCWYKWKRCWMMMAAYG 74
```


Search completed: May 8, 2004, 09:42:36
Job time : 102 secs


```
QY 541 AGACGATATAGTGAGACCTATGGAACCTTACACTATAAATATCACTATTAATTAAGTATGAT 600
DB 4736 AGACGATATAGTGAGACCTATGGAACCTTACACTATAAATATCACTATTAATTAAGTATGAT 4795
QY 601 AAGGGAATATTCAGATATAGTGTACCTCAGTTCAAAAGTGAGCGTCGCGTCGATCTTAAC 660
DB 4796 AAGGGAATATTCAGATATAGTGTACCTCAGTTCAAAAGTGAGCGTCGCGTCGATCTTAAC 4855
QY 661 TTGCGTCCAACTGGTGGGGGCGACATATATGGAAGAAATTCCTGTGATATGCTTTTAT 720
DB 4856 TTGCGTCCAACTGGTGGGGGCGACATATATGGAAGAAATTCCTGTGATATGCTTTTAT 4915
QY 721 GATGATATAGTAACTAACAGAGCTCTTTGGAGATAGATTTTCAAGATTAACATCTTAA 780
DB 4916 GATGATATAGTAACTAACAGAGCTCTTTGGAGATAGATTTTCAAGATTAACATCTTAA 4975
QY 781 TCTGATGGAAATTTTATCTAAGGAAATTAATATGATGACACCAAAAGAAATTTGCATATAC 840
DB 4976 TCTGATGGAAATTTTATCTAAGGAAATTAATATGATGACACCAAAAGAAATTTGCATATAC 5035
QY 841 TTGTCACATCTCTTGGCGGGTAAAGTTTAACTCCAAACAAATGGAACGTCATTAATAT 900
DB 5036 TTGTCACATCTCTTGGCGGGTAAAGTTTAACTCCAAACAAATGGAACGTCATTAATAT 5095
QY 901 GCTGAGCGAGCTCTCTGGAACAACTGGAATAGAAATACAGCTGTCACCATGCCAGAA 960
DB 5096 GCTGAGCGAGCTCTCTGGAACAACTGGAATAGAAATACAGCTGTCACCATGCCAGAA 5155
QY 961 ATCAGTGTTCGGGTGTGTGTGGCTGGACGTTTGAATTTGATGCAAAAGTGGAAAT 1020
DB 5156 ATCAGTGTTCGGGTGTGTGTGGCTGGACGTTTGAATTTGATGCAAAAGTGGAAAT 5215
QY 1021 CCGGAGCTGGACATATATGGTAAATATTAATGTTACTTTACACCAAGTAGTCAAA 1080
DB 5216 CCGGAGCTGGACATATATGGTAAATATTAATGTTACTTTACACCAAGTAGTCAAA 5275
QY 1081 CTCTAG 1086
DB 5276 CTCTAG 5281

RESULT 3
US-10-311-455-1564/c
; Sequence 1564, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1564
; LENGTH: 10279
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1564

Query Match 4.5%; Score 49.4; DB 15; Length 10279;
Best Local Similarity 47.8%; Pred. No. 0.32;
Matches 143; Conservative 0; Mismatches 156; Indels 0; Gaps 0;
```

```
QY 140 AACATAATATTTAATTAACCATATTTACAGCATACAGTCAAGTCTCATATCTCTGTATGATA 199
DB 7605 ATCATATATAAATAATTTAAATAAATAAACAATAAATAAATTAATAATCCAAAAAACA 7546
QY 200 GGATGACATTTTATGTTTGTCTTCTCAATATACACTTTAATGGAGCATCTCCAAACCCAGTG 259
DB 7545 ATAAACCATTTATTAACATTTATTTTCCAAATAAACCAGAAAAAATCTTCTCAATA 7486
QY 260 AGAATCTTAGCAGTTTCATCGGTACGGGTGAACAAATATAACATTTACATTTACCGGAA 319
DB 7485 AAAAAACAATAAATCTTAATTTCAACACGATATAAATAAATAAATAAATAAATAAATAA 7426
QY 320 AAAGAAGTTTATAAATAAAGAGAGCTACAAATTTAAAGGCTATATAAACAATTTCTTCAAAA 379
DB 7425 ACATTAATAATACATATATATATATATATATATATATATATATATATATATATATATAT 7366
QY 380 GTGTTAACTGCGCATCCGGCCCTAACACATTTAACTCAGTCACTTTTAACTGTAATAAACA 438
DB 7365 ACTATATCTTCATAAACCACCAATATTTTATATATATATATATATATATATATATATAT 7307

RESULT 4
US-10-240-589C-86/c
; Sequence 85, Application US/10240589C
; Publication No. US20040076956A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: DNA repair
; FILE REFERENCE: 5013.1008
; CURRENT APPLICATION NUMBER: US/10/240,589C
; PRIOR FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03972
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 148
; SEQ ID NO 96
; LENGTH: 10279
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-589C-86

Query Match 4.5%; Score 49.4; DB 17; Length 10279;
Best Local Similarity 47.8%; Pred. No. 0.32;
Matches 143; Conservative 0; Mismatches 156; Indels 0; Gaps 0;
```

```
QY 380 GTGTTAACTGCCATCCGCGCTTAACACTTAACCTCAGCTCATTTTAACTGTAATAAAAC 438
Db 7365 ACTATATCTTCAATAACACCCCAATATTTATATATCTCTGATTAAATTCATTAAATAC 7307

RESULT 5
US-10-240-453-281/c
; Sequence 281, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BEBLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
; TITLE OF INVENTION: Transcription
; TITLE OF INVENTION: By Means of Assessing the Methylation Status of Genes Associated
; FILE REFERENCE: 5013.1009
; CURRENT APPLICATION NUMBER: US/10/240,453
; PRIOR APPLICATION NUMBER: PCT/EP01/03973
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 281
; LENGTH: 9504
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-281

Query Match 4.4%; Score 48.2; DB 15; Length 9504;
Best Local Similarity 46.8%; Pred. No. 0.61;
Matches 152; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

QY 317 AAAAAGAGTTTAAATAAAAGAGAGCTACAAAATTAAAGGCTATAAAACAATATTGTTC 376
Db 9072 AACATAAATTTCTAAATAATAAAAAATTAACTAAATATAAAAAAATAAAATAAA 9013

QY 377 AAAGTGTTAACTGCCCATCCGCGCTTAACACTTAACCTCAGCTCATTTTAACTGTAATAAA 436
Db 9012 AAATTAACGACAAAAAACAATCTAAACAATTAAATAAAAAATAAAATACAAAAAATA 8953

QY 437 ACGCGGCTTCAGGTCGAAGTTTAAATTTATATATTTCTTCCTGCGGCACTAAAAATTTGC 496
Db 8952 AAACAAAAACGTTTAAATAATTTATATAAAAAAATTATAAAAAACAACACTTAAAAATATA 8893

QY 497 CTTTTCGGTGGATCTGGGATGCTACTCTGAAGTTAAGAGTAAAAAGACGATTAAGTGA 556
Db 8892 TTATAAAAAACAACACTTAATAAATTCATAAATTTAAATTTATCTTTTAAACACTAAAA 8833

QY 557 CCTATGGAAGTACTTAATAATAATACATTAATAATTAACCTGATTAAGGAATATTCAGA 616
Db 8832 AATCACTAAAAAATCAATCTTCTTAAACAATAAATCACTTAACACTTAATAAATA 8773

QY 617 TATGGTTACTCTAGTCAAAAGTGA 641
Db 8772 AATATATCTTCTATTTCTTAAATAAA 8748

RESULT 6
US-10-312-841-2/c
; Sequence 2, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/RO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (379615)
US-10-312-841-2

Query Match 4.4%; Score 48.2; DB 15; Length 3673778;
Best Local Similarity 46.8%; Pred. No. 13;
Matches 152; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

QY 146 ATATATTAATAAACCATATTACAGCATACAGTGAAGTATATCTCTGTATGATAGATGA 205
Db 2363757 ATATATTTAACACACACATACCTAACTAATTTTATATTTTAAATAACAAATAAAAAATCTC 2363698

QY 206 CTTTTTATGTTTGTCTTCTCAATATACACTTAAATCGAGCATGTCACCAACAGTGAGATC 265
Db 2363697 AATTTTAATTAATACTAATATCGAACTCCTAATAATCTACCTACTCTTCCCAAATA 2363638

QY 266 CTAGCAGTTCATCGTCGCGGTAACAAATAATAACATTACAAATTTACGAAAAAGAA 325
Db 2363637 CTAAAAATTACAAAGTAATCGCAACAAACAACTACATTTTTTTTAAATAATAAAAT 2363578

QY 326 GTTTAATAAAAGAGAGAGCTACAAAATTAAAGGCTATAAACAAATTTCTTCAAAAGCTTA 385
Db 2363577 CTTCAAAAAATAAAAAATTTTAAATCTATACAAAAAATAAATAATAATAATAATA 2363518

QY 386 ACTGCCCATCCGCGCTTAACACTTAACCTCAGCTCATTTTAACTGTAATAAAAAACGCGCTT 445
Db 2363517 ATATATAATAAAAAATAAAAAATTTTTCNTTAAATTTTAAAAATAATAATTAATAATA 2363458

QY 446 CAGGTGCAAGTTTATATTTATATATAT 470
Db 2363457 ATATATAATTATATAAACTTATATAT 2363433
```

```
RESULT 7
US-10-424-599-7070
; Sequence 7070, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 7070
; LENGTH: 628
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MBT3847_106392C.1
US-10-424-599-7070
```

```
Query Match 4.1%; Score 44; DB 13; Length 628;
Best Local Similarity 46.4%; Pred. No. 1.6;
Matches 143; Conservative 0; Mismatches 165; Indels 0; Gaps 0;
```


APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 287979
LENGTH: 499
TYPE: DNA
ORGANISM: Human
US-10-027-632-287979

Query Match 4.0%; Score 43.2; DB 16; Length 499;
Best Local Similarity 48.0%; Pred. No. 2.3;
Matches 120; Conservative 1; Mismatches 129; Indels 0; Gaps 0;
QY 669 AACTGCTGGGGGCACATATATGGAGAAATTCCTGATGATGCTTTTATGATGGATA 728
DB 306 AACATTCGAGGAGATAAATGACAAATGCTTTTAAATAGTATTAATGTTATAAA 247
QY 729 TAGTACTAACAGCAGCTCTTTGGAGATAAGATTTCCAGGATAACAATCCATAATCTGATGG 788
DB 246 TAGTAGATAGAAAAGCTCACCTTAATTAATATATATTAATAATTAACCTAAGCTGATTTTCT 187
QY 789 GAAATTTTATCTAAGGAAATTAATGATGACACCAAGAAATTCATATATCTTCTACT 848
DB 186 TGAACCTCAATATGTTTAAACAAAAGCAACAAATATAAATGTTACCATGCCACC 127
QY 849 TCTCTGGCGGTAAAGTTTAACCCAAATGGAACGTCATTAAATATTTGCTGAGCG 908
DB 126 TAGTGGACAGGCTCATTTTGATTTACACCTCGAATCCAGACATTTCAATGCTTCCAA 67
QY 909 AGCTTCTCTG 918
DB 66 AGTGTCTCTG 57

RESULT 11
US-10-027-632-287980/c
Sequence 287980, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 287980
LENGTH: 499
TYPE: DNA
ORGANISM: Human
US-10-027-632-287980

Query Match 4.0%; Score 43.2; DB 16; Length 499;
Best Local Similarity 48.0%; Pred. No. 2.3;
Matches 120; Conservative 1; Mismatches 129; Indels 0; Gaps 0;
QY 669 AACTGCTGGGGGCACATATATGGAGAAATTCCTGATGATGCTTTTATGATGGATA 728
DB 306 AACATTCGAGGAGATAAATGTTACAAATGCTTTTAAATAGTATTAATGTTATAAA 247
QY 729 TAGTACTAACAGCAGCTCTTTGGAGATAAGATTTCCAGGATAACAATCCATAATCTGATGG 788
DB 246 TAGTAGATAGAAAAGCTCACCTTAATTAATATATTAATAATTAACCTAAGCTGATTTTCT 187
QY 789 GAAATTTTATCTAAGGAAATTAATGATGACACCAAGAAATTCATATATCTTCTACT 848
DB 186 TGAACCTCAATATGTTTAAACAAAAGCAACAAATATAAATGTTACCATGCCACC 127
QY 849 TCTCTGGCGGTAAAGTTTAACCCAAATGGAACGTCATTAAATATTTGCTGAGCG 908
DB 126 TAGTGGACAGGCTCATTTTGATTTACACCTCGAATCCAGACATTTCAATGCTTCCAA 67
QY 909 AGCTTCTCTG 918
DB 66 AGTGTCTCTG 57

RESULT 12
US-10-094-240-10/c
Sequence 10, Application US/10094240
Publication No. US20030082637A1
GENERAL INFORMATION:
APPLICANT: ZWIEBEL, LAURENCE J.
TITLE OF INVENTION: ARESTIN GENE, POLYPEPTIDE, AND METHODS OF USE THEREOF
FILE REFERENCE: N8289
CURRENT APPLICATION NUMBER: US/10/094,240
CURRENT FILING DATE: 2001-03-08
PRIOR APPLICATION NUMBER: 10/056,405
PRIOR FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: 60/264,649
PRIOR FILING DATE: 2001-01-26
NUMBER OF SEQ ID NOS: 27
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 10
LENGTH: 4985
TYPE: DNA
ORGANISM: Anopheles gambiae
US-10-094-240-10

Query Match 3.9%; Score 42.8; DB 15; Length 4985;
Best Local Similarity 41.7%; Pred. No. 9.4;
Matches 335; Conservative 0; Mismatches 457; Indels 2; Gaps 1;
QY 8 AGATTTTATTTTATTTTACATTTGTTTTCTCTTCAGTACTTTTACATTTGCTGTAACGG 67
DB 3149 ACAGTTACTGGAATTTACTATTATTTACTTTTATTTTATTTTATTTATGGA 3090
QY 68 CAGATAAATTTCCCGAGATGAAAGCAATTAATTTTTGGCCCGGTGACAGGACG 127
DB 3089 TTACTGGAAATATTAGAGATAAAAAAACAATAATAATAAATAAATAAATAAACA 3030
QY 128 AATCTTCCCCAACATATATATTAATAACCATATTACAGCATACAGTGAAGTCATA 187

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 10:11:11 ; Search time 59 Seconds
(without alignments)
1728.808 Million cell updates/sec

Title: US-09-839-894-10

Perfect score: 1896
Sequence: 1 MNKILPITLFPSSVLPTFA.....EAGQVMGNVITFPSSQTL 361

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq 29Jan04: *
1: Geneseqp1980s: *
2: Geneseqp1990s: *
3: Geneseqp2000s: *
4: Geneseqp2001s: *
5: Geneseqp2002s: *
6: Geneseqp2003as: *
7: Geneseqp2003bs: *
8: Geneseqp2004s: *

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1886	100.0	361	5 AAM50343	ETEC CS4
2	1886	100.0	361	6 ABU09014	Csa opero
3	1881	99.7	1715	6 ABU09015	Csa opero
4	927.5	49.2	364	2 AAY22326	Pilin pro
5	275.5	14.6	359	4 AAB45919	S. enteri
6	120	6.4	1335	6 ABU14929	Protein e
7	119.5	6.3	770	3 AAY50814	D. melano
8	117.5	6.2	1386	6 ABB2573	H. influe
9	117.5	6.2	1449	6 ABB2570	H. influe
10	114	6.0	650	2 AAR47575	Alpha-agg
11	113.5	6.0	513	6 AAB33527	Protein e
12	111.5	5.9	1005	3 AAB01833	Haemophil
13	111.5	5.9	1011	3 AAB01832	Haemophil
14	110.5	5.9	1529	2 AAR41732	High mole
15	110.5	5.9	1601	2 AAW30292	Non-type
16	109	5.8	2383	3 AAB15945	E. coli p
17	109	5.8	2383	6 ABU14928	Protein e
18	108	5.7	2529	6 ABU30910	Protein e
19	107	5.7	1095	3 AAB01835	Haemophil
20	107	5.7	1101	3 AAB01834	Haemophil
21	106.5	5.6	448	6 ABM70719	Photorhab
22	106	5.6	671	2 AAR85290	Streptoco
23	106	5.6	1222	3 AAB01830	H. influe
24	106	5.6	1228	3 AAB01828	Haemophil
25	105	5.6	598	5 ABB47313	Listeria

26 105 5.6 598 6 ABL32687 Protein e
27 104.5 5.5 1440 5 ABE54801 Lactococc
28 103.5 5.5 2399 6 ABU31120 Protein e
29 102 5.4 444 6 AEM69257 Photorhab
30 101 5.4 1338 2 AAR41731
31 101 5.4 1598 2 AAW30291
32 100.5 5.3 1004 3 AAB01841
33 100.5 5.3 1010 3 AAB01840
34 99.5 5.3 969 3 AAB01827
35 99.5 5.3 975 3 AAB01826
36 99.5 5.3 1300 6 ABU39488
37 99 5.2 839 5 ABB56002
38 99 5.2 839 5 ABB98211 Chlamydia
39 99 5.2 839 6 ABU66267 C. psittac
40 99 5.2 932 3 AAB01843
41 99 5.2 998 3 AAB01842
42 99 5.2 1948 6 ADA09347
43 97 5.1 418 3 AAG30907
44 97 5.1 454 3 AAG30906
45 97 5.1 957 3 AAB01839 Haemophil

ALIGNMENTS

RESULT 1
AAM50343 ID AAM50343 standard; protein; 361 AA.
XX
AC AAM50343;
DT 18-FEB-2002 (first entry)
XX
XX ETEC CS4 pilus CsaE tip associated protein.
DE
XX CS4 pilus; enterotoxigenic; ETEC; csa operon; vaccine;
KW diarrhoea; antibacterial; anti-diarrheic.
XX
OS Escherichia coli.
XX
XX Key Location/Qualifiers
FH Peptide 1..23
FT /label= signal_peptide
FT 24..361
FT /label= Mature_protein
XX
PN WO200181582-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US012914.
XX
PR 20-APR-2000; 2000US-0198686P.
XX (UYMA-) UNIV MARYLAND BALTIMORE.
XX
XX Altboum Z, Levine MM, Barry EM;
PI WPI; 2002-049280/06.
XX N-PSDB; AAI70763, AAI70780.
XX
PT New nucleotide sequence, useful as immunogenic agent for generating
PT immune response against recombinant product of the operon, comprises csa
PT operon which encodes enterotoxigenic Escherichia coli-CS4 pill.
XX
PS Claim 10; Page 59; 81pp; English.
XX
CC The present sequence is that of the tip associated protein CsaE of
CC enterotoxigenic Escherichia coli (ETEC) strain E11891A. CsaE is encoded
CC by the csaE gene (see AAI70763) of the E. coli E11891A csa operon. This
CC operon includes 5 contiguous genes, csaA-csaE, which encode the synthesis
CC of ETEC-CS4 pill. It has been expressed in attenuated Shigella strain
CC CVD1204 guabA, constructing the Shigella expressing CS4 fimbriae vaccine

CC strain CVD1204 (pCA2-CS4). The CsaE protein has a calculated mol.wt. of
 CC 40102.4 and a theoretical pI of 8.74. It shows homology to similar
 CC proteins from other ETEC fimbriae. Recombinant CsaA-CsaE polypeptides are
 CC used in claimed immunogenic compositions to generate an immune response
 CC in a subject. these prevent ETEC colonisation, and hence protect against
 CC diarrhoea
 XX
 SQ Sequence 361 AA;

Query Match 100.0%; Score 1886; DB 5; Length 361;
 Best Local Similarity 100.0%; Pred. No. 1.6e-168;
 Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNKILFIPTLFFSSVLFTFAVSADKIPGDESITNIFGPRDRNESSPKHNILNNHITAYSE 60
 DB 1 MNKILFIPTLFFSSVLFTFAVSADKIPGDESITNIFGPRDRNESSPKHNILNNHITAYSE 60

QY 61 SHLYDRMTFLCSSHNTLNGACPTSENPPSSSVSGETNITLQFTKRSLIKRELQIKGY 120
 DB 61 SHLYDRMTFLCSSHNTLNGACPTSENPPSSSVSGETNITLQFTKRSLIKRELQIKGY 120

QY 121 KOLLFKSVNCPGSLTNSAHFNCKNAASGASLYLYIPAGELKNLPFGGIWDATLKLVRK 180
 DB 121 KOLLFKSVNCPGSLTNSAHFNCKNAASGASLYLYIPAGELKNLPFGGIWDATLKLVRK 180

QY 181 RRYSETYGTVTINITIKLTDKGNIQIWLPOFKSDARVDLNRPTGGTYIGRNSVDMCFY 240
 DB 181 RRYSETYGTVTINITIKLTDKGNIQIWLPOFKSDARVDLNRPTGGTYIGRNSVDMCFY 240

QY 241 DGYSTNSSSLEIRFQDNNPKSDGKFYLRKINDTKEIAYTSLLLAGKSLTPNGTSLNI 300
 DB 241 DGYSTNSSSLEIRFQDNNPKSDGKFYLRKINDTKEIAYTSLLLAGKSLTPNGTSLNI 300

QY 301 ADAASLETNNRITAVTMPPEISVPVLCWPGRLQDAXVENPEAGQYMGNIIVTFPSSQT 360
 DB 301 ADAASLETNNRITAVTMPPEISVPVLCWPGRLQDAXVENPEAGQYMGNIIVTFPSSQT 360

QY 361 L 361
 DB 361 L 361

RESULT 2
 ABU09014
 ID ABU09014 standard; protein; 361 AA.
 AC ABU09014;
 XX
 DT 16-JUN-2003 (first entry)
 DE Csa operon recombinant product, CsaE.
 XX
 KW CsaE; csa operon; CS4 pilus; bacterial pili protein; ETEC; CS4;
 KW enterotoxigenic Escherichia coli; multivalent Shigella-ETEC; diarrhoea;
 KW Shigella; antibacterial.
 XX
 OS Escherichia coli.
 XX
 PN US2002178968-A1.
 XX
 PD 28-NOV-2002.
 XX
 PF 20-APR-2001; 2001US-00839894.
 XX
 PR 20-APR-2000; 2000US-0198626P.
 XX
 PA (ALTB/) ALTBOUM Z.
 PA (LEVI/) LEVINE M M.
 PA (BARR/) BARRY E M.
 XX
 PI Altboum Z, Levine Mx, Barry EX;
 XX
 DR WPI; 2003-352604/33.

DR N-PSDB; ABX93757.
 XX
 PT Novel isolated nucleic acid comprising csa operon encoding proteins
 PT required for producing CS4 pili, useful for generating immune response in
 PT vertebrate against the enterotoxigenic Escherichia coli.
 XX
 PS Claim 10; Page 31-32; 58pp; English.

XX The invention relates to an isolated nucleotide sequence comprising a csa
 CC operon (encoding proteins required for producing CS4 pili) or its
 CC functional fragment. An immunogenic composition comprising a recombinant
 CC product of a csa operon and a carrier, is useful for generating an immune
 CC response in a subject, which involves contacting the subject with the
 CC immunogenic composition. The recombinant product of the composition is
 CC the CS4 antigen (bacterial pili protein) and is provided in an acellular
 CC or cellular composition. The nucleic acid is useful for producing a
 CC polypeptide product from a csa operon or functional fragment, which
 CC involves providing a nucleic acid, introducing the csa operon in an
 CC expression vector, such that a recombinant host cell is produced and
 CC subjecting the recombinant host cell to conditions such that a protein
 CC from the csa operon is expressed. The nucleic acid encoding at least an
 CC immunogenic portion of the csa operon or a polypeptide encoded by the
 CC nucleic acid is useful for generating an immune response in a vertebrate
 CC against ETEC (enterotoxigenic Escherichia coli). The nucleic acid is also
 CC useful for identifying polynucleotides encoding other proteins with
 CC biological functions similar to that of the csa operon and for creating a
 CC multivalent Shigella-ETEC immunogenic composition that will protect from
 CC diarrhoea caused by either Shigella or CS4 expressing ETEC strains. This
 CC sequence represents a recombinant product of a csa operon, CsaE
 XX Sequence 361 AA;

Query Match 100.0%; Score 1886; DB 6; Length 361;
 Best Local Similarity 100.0%; Pred. No. 1.6e-168;
 Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNKILFIPTLFFSSVLFTFAVSADKIPGDESITNIFGPRDRNESSPKHNILNNHITAYSE 60
 DB 1 MNKILFIPTLFFSSVLFTFAVSADKIPGDESITNIFGPRDRNESSPKHNILNNHITAYSE 60

QY 61 SHLYDRMTFLCSSHNTLNGACPTSENPPSSSVSGETNITLQFTKRSLIKRELQIKGY 120
 DB 61 SHLYDRMTFLCSSHNTLNGACPTSENPPSSSVSGETNITLQFTKRSLIKRELQIKGY 120

QY 121 KOLLFKSVNCPGSLTNSAHFNCKNAASGASLYLYIPAGELKNLPFGGIWDATLKLVRK 180
 DB 121 KOLLFKSVNCPGSLTNSAHFNCKNAASGASLYLYIPAGELKNLPFGGIWDATLKLVRK 180

QY 181 RRYSETYGTVTINITIKLTDKGNIQIWLPOFKSDARVDLNRPTGGTYIGRNSVDMCFY 240
 DB 181 RRYSETYGTVTINITIKLTDKGNIQIWLPOFKSDARVDLNRPTGGTYIGRNSVDMCFY 240

QY 241 DGYSTNSSSLEIRFQDNNPKSDGKFYLRKINDTKEIAYTSLLLAGKSLTPNGTSLNI 300
 DB 241 DGYSTNSSSLEIRFQDNNPKSDGKFYLRKINDTKEIAYTSLLLAGKSLTPNGTSLNI 300

QY 301 ADAASLETNNRITAVTMPPEISVPVLCWPGRLQDAXVENPEAGQYMGNIIVTFPSSQT 360
 DB 301 ADAASLETNNRITAVTMPPEISVPVLCWPGRLQDAXVENPEAGQYMGNIIVTFPSSQT 360

QY 361 L 361
 DB 361 L 361

RESULT 3
 ABU09015
 ID ABU09015 standard; protein; 1715 AA.
 XX
 AC ABU09015;
 XX
 DT 16-JUN-2003 (first entry)
 XX

Csa operon related protein.

Csa operon; CS4 pilus; bacterial pili protein; ETEC; CS4;
enterotoxigenic Escherichia coli; multivalent Shigella-ETEC; diarrhoea;
Shigella; antibacterial.

Escherichia coli.

US2002176868-A1.

28-NCV-2002.

20-APR-2001; 2001US-00839894.

20-APR-2000; 2000US-0198626P.

(ALTE/) ALTBOUM Z.

(LEVI/) LEVINE M.M.

(BARR/) BARRY E.M.

Altboum Z, Levine MM, Barry EM;

WPI; 2003-352604/33.

Novel isolated nucleic acid comprising csa operon encoding proteins
required for producing CS4 pili, useful for generating immune response in
vertebrate against the enterotoxigenic Escherichia coli.

Disclosure; Page 35-41; 58pp; English.

The invention relates to an isolated nucleotide sequence comprising a csa
operon (encoding proteins required for producing CS4 pili) or its
functional fragment. An immunogenic composition comprising a recombinant
product of a csa operon and a carrier, is useful for generating an immune
response in a subject, which involves contacting the subject with the
immunogenic composition. The recombinant product of the composition is
the CS4 antigen (bacterial pili protein) and is provided in an acellular
or cellular composition. The nucleic acid is useful for producing a
polypeptide product from a csa operon or functional fragment, which
involves providing a nucleic acid, introducing the csa operon in an
expression vector, such that a recombinant host cell is produced and
subjecting the recombinant host cell to conditions such that a protein
from the csa operon is expressed. The nucleic acid encoding at least an
immunogenic portion of the csa operon or a polypeptide encoded by the
nucleic acid is useful for generating an immune response in a vertebrate
against ETEC (enterotoxigenic Escherichia coli). The nucleic acid is also
useful for identifying polynucleotides encoding other proteins with
biological functions similar to that of the csa operon and for creating a
multivalent Shigella-ETEC immunogenic composition that will protect from
diarrhoea caused by either Shigella or CS4 expressing ETEC strains. This
sequence represents a csa operon related protein of the invention

Sequence 1715 AA;

Query Match 99.7%; Score 1881; DB 6; Length 1715;

Best Local Similarity 100.0%; Pred. No. 4.7e-167;
Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 NKILFIPTLFFSSVLTFFAVSADKIPGDESITNIFGPRDNSSPKHNILNHITAYSS 61
|||||
1264 NKILFIPTLFFSSVLTFFAVSADKIPGDESITNIFGPRDNSSPKHNILNHITAYSS 1323

62 HTLYDRMTFLCLSSHNTLNGACPTSENPPSSSVSGETNITLQTEKRSLIKRELQIKGYK 121

1324 HTLYDRMTFLCLSSHNTLNGACPTSENPPSSSVSGETNITLQTEKRSLIKRELQIKGYK 1383

122 QILFKVNCPSGTLNSAHFNCKNKAASGASLYLYIPAGELKNLPFGGIWDATLKLVRKR 181

1384 QILFKVNCPSGTLNSAHFNCKNKAASGASLYLYIPAGELKNLPFGGIWDATLKLVRKR 1443

182 RYSETYGTYTINIIKLTGKNIQIWLPOFKSDARVDLNLRTGGGTYIGRNSVDVCFYD 241

1444 RYSETYGTYTINIIKLTGKNIQIWLPOFKSDARVDLNLRTGGGTYIGRNSVDVCFYD 1503

QY 242 CYSTNSSLEIRFQDNPKSD3KEVLRKINDTKETIAVITLSLLLAGKSLTPNGTSLNIA 301
|||||
DB 1504 CYSTNSSLEIRFQDNPKSD3KEVLRKINDTKETIAVITLSLLLAGKSLTPNGTSLNIA 1563
|||||
QY 303 DAASLETNNRITAVTMPEISVPVLCWPEGR-QLDAKVENPEAGQTMGNINVTFTPSQTL 361
|||||
DB 1564 DAASLETNNRITAVTMPEISVPVLCWPEGR-QLDAKVENPEAGQTMGNINVTFTPSQTL 1623
|||||

RESULT 4

AAV22326

ID AAY22326 standard; protein; 364 AA.

XX AAY22326;

AC AAY22326;

DT 22-SEP-1999 (first entry)

DE Pili protein CotD.

XX CS2 gene cluster; CotA; CotB; CotC; CotD; pili protein; immunogen;

KW enterotoxigenic E. coli; human upper intestine; diarrhoeal disease;

KW enteric infection; therapy.

XX Escherichia coli.

OS US5932715-A.

XX 03-AUG-1999.

PD 07-JUN-1995; 95US-00483101.

XX 07-JUN-1995; 95US-00483101.

PR (UYEM-) UNTV EMORY.

XX Caron J, Froehlich B, Scott JR;

XX WPI; 1999-443623/37.

DR N-PSDB; AAX84848.

XX Isolated nucleic acids encoding Escherichia coli CS2 pili proteins

PT useful for vaccinating against diarrheal diseases caused by Escherichia

PT coli.

XX Claim 3; Col 45-48; 35pp; English.

XX This sequence represents a CS2 pili protein of the invention, encoded by

CC (which also encodes CotA, CotC, and CotB). CS2 pili are long

CC proteinaceous molecules thought to mediate attachment of enterotoxigenic

CC E. coli (ETEC) to and/or promote colonisation of the human upper

CC intestine. The CS2 gene cluster may be used to produce immunogens for

CC vaccinating patients against diarrhoeal diseases caused by ETEC bacteria.

CC This type of enteric infection is a major cause of death among infants in

CC developing countries and in immunocompromised (e.g. Acquired Immune

CC Deficiency Syndrome (AIDS)) or elderly adults. The vaccine comprises more

CC than 1 antigenic determinant (epitopes) from more than 1 pilus type to be

CC effective against more than 1 type of ETEC infection

XX Sequence 364 AA;

SQ Query Match 49.2%; Score 927.5; DB 2; Length 364;

Best Local Similarity 48.1%; Pred. No. 2.4e-78;

Matches 176; Conservative 65; Mismatches 118; Indels 7; Gaps 3;

DE

XX

KW

KW

KW

XX

OS

XX

FN

XX

XX

PD

XX

PF

XX

XX

PR

XX

PA

PA

PA

XX

XX

PI

XX

XX

DR

XX

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

XX

SQ

Query Match

Best Local Similarity

Matches 360;

Conservative

0; Mismatches

0; Indels

0; Gaps

0;

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 GC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from Wipo at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 1335 AA;

Query Match 6.4%; Score 120; DB 6; Length 1335;
 Best Local Similarity 20.4%; Pred. No. 0.15;
 Matches 80; Conservative 66; Mismatches 156; Indels 90; Gaps 15;
 QY 13 SSVLTFEFAVADKIPGDESITNIFGRDRNESPKNHILNHTAY---SESHLYDRMT 69
 DB 746 SAKIATLSASNGVLANEAMNTVSVNVVADEGS---NPINDHTVFAVLGSGATSFNQON 802
 QY 70 FLCLSHNTLNGACPTSENPFSSSVSGETNITLOFTEKRSLI-----K 112
 DB 803 ---FAKTDVNGLA-TFDLKSQKQEDNTVEVLNGVKQTLIVSFVGDSSSTAQVDLQSK 857
 QY 113 RELQIKGY-KQLLFKSVNCPGLTNSAHFNCNKAASGASLYIYPAGELKNLPFGGIW 171
 DB 858 NSVADGNDSATMTATVRDAKGNLLNDVKTFTNVNSAAKLSQTEVNSHD-----GTA 910
 QY 172 DATLKLVRKRYSETGYTITNITIKLTPKGNIQIWLPOFKSDARVDLNLRTGGTYIG 231
 DB 911 TAILT-----SLKNGDYTTVASVSSGSAQNOQVIFIGDQTAALTLSV-PSGDTIV-- 960
 QY 232 RNSVDMCFYDGYSTNSSLE----IRFQNNPKSGDKFYLRKINDDTKEIAYTLLLAG 287
 DB 961 -----TNTAPLHMTATLQCKXGNPLKDKKEITFSPND-----VASRFSISNSG 1003
 QY 288 KSLTFTNGTSL-----NIADAASLETWNRTAVTMTPEISVPLVC 327
 DB 1004 KGMTDSNGTAIASLQTLGTHMTATRLANSVSDTQPMTFVADKRAVYVLTQSKAEII 1063
 QY 328 WFG--RLQDQAKVENPEAGQYGNINVTTPS 357
 DB 1064 GNGVDETTITATVKDP-FDNVKNLSVVFATS 1094

RESULT 7

AAV50814

ID AAV50814 standard; protein; 770 AA.

XX AAV50814;

AC AAV50814;

XX 17-FEB-2000 (first entry)

DT 17-FEB-2000 (first entry)

XX

DE D. melanogaster acetyl-cho-line receptor protein from clone Da7.
 XX Acetyl-cho-line receptor; nicotinic; insect; insecticide; screening;
 KW neurotransmission; plant protection agent; conductance; Achr.
 XX Drosophila melanogaster.
 OS
 XX DE19819829-A1.
 XX 11-NOV-1999.
 PD
 XX 04-MAY-1998; 98DE-010-9829.
 PF
 XX 04-MAY-1998; 98DE-010-9829.
 PR
 XX (FARB) BAYER AG.
 PA Adamczewski M, Oellers N, Schulte T;
 XX
 PI WPI: 2000-014207/02.
 XX N-PSDB; AA224475.
 DR
 XX New nucleic acid encoding a nicotinic acetylcholine receptor from
 PT insects, used to identify potential insecticides.
 XX Example 1a; Page 12-14; 26pp; German.
 PS
 XX This invention describes a novel nucleic acid (NA) encoding a nicotinic
 CC acetyl-cho-line receptor (I) from insects which can be used as an
 CC insecticide. Inhibitors of (I) interfere with neurotransmission. (I)
 CC (also vectors containing it, its regulatory regions, and antibodies
 CC directed against (I)-encoded proteins) are used to screen for: (a) plant
 CC protection agents that alter conductance of ACHR, potentially useful as
 CC insecticides, or (b) genes which encode polypeptides that are involved in
 CC formation of functionally related ACHR in insects. (I) are also used to
 CC isolate and characterize the specified regulatory regions and for
 CC recombinant production of (II). This sequence represents an acetyl-
 CC choline receptor isolated from *Drosophila melanogaster*
 XX
 SQ Sequence 770 AA;

Query Match 6.3%; Score 119.5; DB 3; Length 770;
 Best Local Similarity 20.4%; Pred. No. 0.074;
 Matches 33; Conservative 48; Mismatches 140; Indels 135; Gaps 20;
 QY 3 ILFPSSVLETFVAADKIPGDESITNIFGRDRNESPKNHILNHTAY---SESHLYDRMT 54
 DB 243 TIATFISYLGSAQAQKNSGSSSSSN---SSNNSSTQILNGLNKHSWIFLLIYLNLGAS 304
 QY 55 --ITAVSBSHTLYDRMTFLCLSSHTLNGACPTSENPFSSSVSGETNITLOFTEKRSLIK 112
 DB 305 VCIAGYHEKLLHD-----LLDPYTLERFVNESDPLQSLFGLTLMQIYDVERKNQILV 359
 QY 113 RELQIKGYKQLLFKSVNCPGLTNSAHFNCNKAASGASLYIYPAGELKNLPFGGIW 172
 DB 360 TNVWLK-----LEWDMNLRWNTSDYGVVK-----DLR-IPSHRIWK 395
 QY 173 ATLKLVRKRYSET---GYTITNITIKLTPKGNIQIWLPO--FKSDARVDLNLRTGG 226
 DB 395 PDVLM-----YNSADEGFDGTQTNVVR--NNGSC-LYVPEGIFKS-CKIDITWFP--- 444
 QY 227 GTYIGKNSVDMCF-----YDGYSTNSSSLEIRFQDNNPKS-----DGKPYLRKINDDTKE 276
 DB 445 ---FDDQRCMAFGSWTYDGF-----QLDLQDDEGDISSYVINGEWELLGVGKRE 496
 QY 277 IAYT-----SLLIAGKSLTPTNGTSLNIADA 303
 DB 497 IYVNCPEPYIDITAIIRRTLYFFNLIITPCVLIASYALLGFTLPDSCSEKLSGVT 556
 QY 304 ASLE-TNNWRITAVTMPEIS--VPV-----LCW-----PGR 331
 DB 557 ILLSLTVFLNVAETMPATSDAIVLIRIVFLCWLPLWILMRSPGR 602

Db	850	NRGTSINIKQGVVKIQGDITNNGNLMTTNSVNOQTIIINGNITNKKGDLNKKIKANA	909
Qy	216	RVDL--NLRPTGGTYIGRNSVDMCMFYDGYSTNSSSLEIRFQDNNPKSD-GKFFVLRKIND	272
Db	910	BIQIGGNISQKEGNLTISSDKINI-----TKRIEKADTQCGNSDGVASNANLTI	950
Qy	273	DTKETAYTSLSLLAG--KSLTPTPTGSLNITADAASLETNWRITAVTTPMPEISVFLCWP	329
Db	951	TKKETLTLDNLNIGFNKABITAKDNDLIIIGKASSDNSNAKQTTPDKVDKSLIS--AGN	1018
Qy	330	GRLOLDAKVERPEA-----QYMGNNINVTTPSSQTL	361
Db	1019	HNVTLSKSVTSNSDGSFGNSDDNNIGLTTSAKDVTV	1056
RESULT 9			
ID	ABB82570	standard, protein; 1449 AA.	
XX	ABB82570;		
AC	ABB82570;		
DT	04-FEB-2003	(first entry)	
XX	DE	H. influenzae BASB223 polypeptide.	
XX	KW	BASB223; BASB224; high molecular weight protein A; immunostimulant;	
KW	KW	antibacterial; vaccine; medicament; bacterium.	
XX	OS	Haemophilus influenzae.	
XX	PN	WO200279237-A2.	
XX	PD	10-OCT-2002.	
XX	PF	12-MAR-2002; 2002WO-EP003210.	
XX	PR	13-MAR-2001; 2001GB-00006155.	
PR	N-PSDE	2001GB-00006155.	
XX	PA	(GLAX) GLAXOSMITHKLINE BIOLOGICALS SA.	
XX	PI	Thomnard J;	
XX	PI	WPI; 2003-040650/03.	
DR	N-PSDE	ABV74994.	
XX	XX	New BASB223 and BASB224 polypeptides and polynucleotides, useful for	
PT	PT	preparing a medicament for generating an immune response in an animal, or	
PT	PT	as research reagents and materials for discovering treatments for human	
PT	PT	diseases.	
XX	XX	Claim 3; Page 67-68; 109pp; English.	
XX	XX	The invention relates to novel polypeptides BASB223 and BASB224, isolated	
CC	CC	from non-typeable H. influenzae (NTHi). The polypeptides are related to	
CC	CC	H. influenzae high molecular weight protein A (Hmwa). Compositions	
CC	CC	comprising the BASB223 and BASB224 polypeptides and polynucleotides are	
CC	CC	useful for preparing a medicament for generating an immune response in an	
CC	CC	animal. The BASB223 and BASB224 polypeptides and polynucleotides may be	
CC	CC	used as research reagents and materials for discovery of treatments and	
CC	CC	diagnosis for diseases, particularly human diseases, and as immunogens	
CC	CC	for producing antibodies immunospecific the such polypeptides and	
CC	CC	polynucleotides. The polynucleotides may be used in the discovery and	
CC	CC	development of antibacterial compounds, in genetic immunization, in the	
CC	CC	development of prophylactic or therapeutic agents for bacterial	
CC	CC	infection, particularly non-typeable H. influenzae. The present sequence	
CC	CC	represents the BASB223 or BASB224 polypeptide partial sequence	
XX	XX	Sequence 1386 AA;	
XX	XX	Query Match 6.2%; Score 117.5; DB 6; Length 1386;	
XX	XX	Best Local Similarity 20.4%; Pred. No. 0.27;	
XX	XX	Matches 81; Conservative 62; Mismatches 136; Indels 119; Gaps 17;	
Qy	45	SPKHNLN---HITAYSESHTLYDRMTPLCLSSHTNLGACPTGENPSSSVSGET---	98
Db	697	SSRINTNGSSLSITGDMPAKKIFDKNDLV-NA-----TNSVSIIEVEGTDTKL	747
Qy	99	-----NITQTEKSLIKRELQIKGYKQLLFKSNVCNCSGTLNSAHFNC	143
Db	748	EYGLVADGNITVEGGNVLGSKAKTHITKNVSVXS-----NANLTSSANPNV	796
Qy	144	NKAAS-GASLIYV-PAGEIKNLPFGGIW-----DATLKLVRKRYSTYGTYT	191
Db	797	HKGALTIGGS-----ANTQGNLTANGDTVEVAGDVIVSDDAKFAETKNLNTI-GTFT	849
Qy	192	IN-----ITTK-----LTDKGNIOIW-----LPOFKSDA	215

Matches 81; Conservative 62; Mismatches 136; Indels 119; Gaps 17;

QY 45 SPKHLNKK---HITAYSESHLYDRMTFLCLSSHTNLNGACPTSPSSSVSGPT--- 98

Db 760 SSRINVTGSS--SITGDMPAKKIFDIKNDLVINA-----TNSNVGITEVEGTDTKL 810

QY 99 -----NITQFTKRSLLKRELQIKGYKQKLLPKSVNCPGSLTNSAHENC 143

Db 811 EYGLVADGNITVEGGNVTLGSKAKTHITKNSVK-----NANLTSSANFV 859

QY 144 NKAAAS-GASLYLYIPAGELKNLPFGGIW-----DATLKLVRKRYSETYGT 191

Db 860 HKGALTIGGS-----ANIQGNLTANGDTVEVAGDVIVSDDAKFAETKNLNIT-GTFT 912

QY 192 IN-----ITIK-----LTKGNITQIW-----LPOPKSDA 215

Db 913 NNGTSEINIKQGVVKLQGDITNNGN.NITTNASVNOKTIINGNITNKKGLNLIKDKANA 972

QY 216 RVDL--NLRPTGGGTIGRNSVDMCFYDGYSTNSSSLEIRFQDNNPKSD-GKPYLRKIND 272

Db 973 ETQIGNISQKGNLTISSDKINI-----TKRIKADTDQNSDSGVASNANLTI 1023

QY 273 DTKEIAYTLSSLAG---KSLPTNGTSLNIADAASLETNMRITAVTTPETISVPVLCWP 329

Db 1024 KTKELTLTDNLNIGSPNKAETAKDNSDLIGKASSDNSNAKQITFDKVKDSKIS--ACN 1081

QY 330 GLQLDAKVENDEA-----QVMGNINVTFTPSOTL 361

Db 1082 HNVTLNSKVTESDGSSTGSGDDNIGHTISAKDVTV 1119

RESULT 10

AAR47575

ID AAR47575 standard; protein; 650 AA.

XX AAR47575;

AC AAR47575;

XX AAR47575;

DT 25-MAR-2003 (revised)

DT 19-JUL-1994 (first entry)

XX

XX Alpha-agglutinin of Saccharomyces cerevisiae.

DE Immobilisation; enzyme; cell wall; alpha agglutinin; AGA 1; FLO 1;

KW Major cell wall protein; glycosyl-phosphatidyl-inositol;

KW anchoring protein; alpha factor; alpha-agglutinin; invertase; inulinase;

KW alpha-amylase; Saccharomyces cerevisiae; enzymatic process; fermentation;

KW biodegradation; catalysis.

XX Saccharomyces cerevisiae.

CS

XX WO9401567-A1.

PN

XX 20-JAN-1994.

XX

XX 07-JUL-1993; 93WO-EP001763.

XX

PR 08-JUL-1992; 92EP-00202080.

PR 14-DEC-1992; 92EP-00203899.

XX

XX (UNIL) UNILEVER PLC.

PA (UNIL) UNILEVER NV.

XX

XX Klis FM, Schreuder MP, Teschka H, Verrips CT;

PI

XX WPI; 1994-035071/04.

DR N-PSDB; AAQ54012.

XX

XX Immobilisation of enzymes to microbial cell wall - by prodn. of fusion

PT protein of enzyme linked to anchoring protein.

XX

XX Example 1; Page 32-39; 99pp; English.

PS

XX The alpha-agglutinin is used in a method to immobilise enzymes to a

microbial cell wall. The coding sequence is used in the production of a recombinant polynucleotide which comprises a structural gene encoding a protein with catalytic activity and at least part of a gene encoding at least the C-terminus of a protein capable of anchoring in a eukaryotic or prokaryotic cell wall. The anchoring fragment or protein is selected from alpha agglutinin, AGA 1, FLO 1, major cell wall protein of lower eukaryotes or a proteinase of lactic acid bacteria. The recombinant polynucleotide preferably also comprises a sequence encoding a signal peptide to ensure secretion of the expressed product. The signal peptide is preferably derived from glycosyl-phosphatidyl-inositol, anchoring protein, alpha factor, alpha-agglutinin, invertase or inulinase, alpha-amylase of Bacillus or proteinases of lactic acid bacteria. The host microorganism can be used for performing enzymatic processes on an industrial scale. (Updated on 25-MAR-2003 to correct PN field.)

XX

SQ Sequence 650 AA;

Query Match 6.0%; Score 114; DB 2; Length 650;

Best Local Similarity 20.9%; Pred. No. 0.19;

Matches 98; Conservative 52; Mismatches 147; Indels 172; Gaps 23;

QY 4 ILFIETLFFSSVL-----VSADKIPG-----DESTNIFGPRDRNE--- 43

Db 3 ILWLFLSALASAININDITFSNLBITPLTANKPQDQGTATDFSLADASSREGDEFTL 67

QY 44 SSP---KNNILNNHITA-----YSESHLYDRMTFLC-----LSSHNTLN 80

Db 63 SMPHVYRIKLLNSSQTATISLADGTEAFKCVYVQQAALYEVNTTCTAQNLDSSVNTID 127

QY 81 GACPTSENPSSSVSGETNITLQFTKRSLLKRELQIKGYKQKLLPKSVNC-PSGLTNSA 139

Db 123 GSITFSLNPFSGGSSYEVEL-----ENAKFFKSGPMLVKLGNDSDVNVDFEAAFTENVF 182

QY 143 HENCNKAASGASLYL--YIPAGELKNLPFGGIWDATLKLVRKRYSETYGTYNITIK 197

Db 183 HSGRSTGYGSFESVHLGYCPNGYF---LGG---TEKI-----DYSSNNNVDL- 225

QY 193 LTDKNIQI-----WLPQKSDARVDNLNRPFGGTIGRN---SVMCFYDG--- 242

Db 226 --DCSSVQVYSSNDFNDWFFQSYNDTNADV-----TCFGSNLWITLDBKLYDGEML 275

QY 243 -----YSTNSSLE--IRFQDNV--PKSDGKF 265

Db 276 WNALQSLPANVNTIDHALEFYTCLDTIANTTYATQSTTREFIVYQGRNLGTASAKSS 335

QY 265 YLRKINDTKBI---AYTLS-----ILLAGKSLTPTNGTSLNIADAA 304

Db 336 PISTVTTDLTSLNTSAYSTGISTVETGRTTSEVISHVVTSTKLSPTATSLTIAQTS 395

QY 305 SLETNWN-----RITAVTMPE-SVPVLCWPGRL 332

Db 396 IYSTDSNITVGTDIHTTSEVISDVETISRETASTVVAAPTSTGTGTGAM 444

RESULT 11

ABU33527

ID ABU33527 standard; protein; 513 AA.

XX ABU33527;

AC ABU33527;

XX

XX 19-JUN-2003 (first entry)

XX

DE Protein encoded by prokaryotic essential gene #19054.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX Legionella pneumophila.

OS

XX WO20027183-A2.

PN

XX 03-OCT-2002.

XX

XX 21-MAR-2002; 2002WO-JSC09107.

[illegible]

Db 210 -----GRVLSFSDIENLSGKSSVTEGQLQIQNILN_QFALKKINTTLTKHKDIOF 263

QY 280 ----TSL-----LLAGKSLTPNG-----TSLNIADAASLETNWRITA 315

Db 264 NPLTISLVNGESIGMDYVIATQQLINQATNLDGKQLITSLKHPAISGNLDYSIHAS 323

QY 316 VTMPISIVPLCWPRLOL_DAKVENPEAGQYMGINV-----TTFPSS 358

Db 324 IFLKALSIESLSKGTICLKGEVYNINLDQLNNKLVKLSLMTETPDN 373

RESULT 12

AA01833

ID AA01833 standard; protein; 1005 AA.

XX

AC AA01833;

XX

DT 12-SEP-2003 (revised)

DT 11-SEP-2000 (first entry)

XX

DE Haemophilus influenzae strain K21 mature HMW2A protein, SEQ ID NO:41.

XX

KW Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;

KW non-typhable Haemophilus influenzae; NTHi; non-encapsulated;

KW recombinant production; Escherichia coli; anti-bacterial; vaccine;

KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;

KW detection; diagnosis.

XX

XX Haemophilus influenzae; strain K21.

OS

XX W0200020609-A2.

PN

XX 13-APR-2000.

PD

XX 07-OCT-1999; 99WO-CA000938.

PF

XX 07-OCT-1998; 98US-00157568.

PR

PR 08-DEC-1998; 98US-00206942.

XX

XX (CONK-) CONNAUGHT LAB LTD.

PA

XX Loonsmore SM, Yang Y, Klein MH;

PI

XX WPI; 2000-303789/26.

DR

DR N-PSDB; AAA52182.

XX

XX Nucleic acid molecule for producing recombinant high molecular weight

PT proteins of Haemophilus which are used as a vaccine to provide protection

PT against Haemophilus induced diseases in humans.

XX

XX Claim 8; Fig 21A-O; 307pp; English.

XX

XX The invention relates to the recombinant production of Haemophilus

CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The

CC expression construct used to effect recombinant expression comprises a

CC promoter functional in E. coli (e.g., the T7 promoter) operably linked to

CC a modified hmwABC operon from a non-typhable (non-encapsulated) H.

CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene

CC clusters termed hmwAABC and hmw2ABC. Each hmwABC operon comprises hmwA,

CC hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins,

CC and the hmwB and hmwC genes encode accessory proteins which are

CC responsible for post-translational processing and secretion of the HMWA

CC proteins. The modified hmwABC operon used in the expression construct of

CC the invention contains an A gene modified such that it encodes only the

CC mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)

CC and HMWA proteins (AA01824-B01845) from the non-typhable H. influenzae

CC strains Joyce, Kl, K21, LCD02, PMH1, 15 and 12. The nucleic acids and

CC vectors are used for the production of recombinant H. influenzae HMW

CC proteins which can be used as vaccines to mediate a humoral or cell-

CC mediated immune response to provide protection against diseases in humans

CC caused by H. influenzae (e.g., otitis media, epiglottitis, pneumonia and

CC tracheobronchitis). The HMW proteins are also useful as antigens in

CC immunoassays for detecting antibodies against Haemophilus HMW proteins

PT and/or HmW peptides. The nucleotide sequences encoding the HmW proteins
PT can be used to isolate and clone hmw genes from other non-typeable
XX strains of Haemophilus via hybridisation reactions. The present sequence
PS represents a mature HmW protein from a non-typeable strain of H.
XX influenzae. (Updated on 12-SEP-2003 to standardise OS field)

XX
SQ Sequence 1005 AA;
Query Match 5.9%; Score 111.5; DB 3; Length 1005;
Best Local Similarity 20.3%; Pred. No. 0.62;
Matches 68; Conservative 65; Mismatches 147; Indels 55; Gaps 14;

CC The invention relates to the recombinant production of Haemophilus
CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC expression construct used to effect recombinant expression comprises a
CC promoter functional in E. coli (e.g., the T7 promoter) operably linked to
CC a modified hmwABC operon from a non-typeable NTHi (non-encapsulated) H.
CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC hmwB and hmwC genes. The hmwA genes encode the structural HMW proteins
CC and the hmwB and hmwC genes encode accessory proteins which are
CC responsible for post-translational processing and secretion of the HMW
CC proteins. The modified hmwABC operon used in the expression construct of
CC the invention contains an A gene modified such that it encodes only the
CC mature HMWA. The invention also discloses hmwA genes (AAAS2175-AS2198)
CC and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC strains Jovc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC vectors are used for the production of recombinant H. influenzae HMW
CC proteins which can be used as vaccines to mediate a humoral or cell-
CC mediated immune response to provide protection against diseases in humans
CC caused by H. influenzae (e.g., otitis media, epiglottitis, pneumonia and
CC tracheobronchitis). The HMW proteins are also useful as antigens in
CC immunoassays for detecting antibodies against Haemophilus, HMW proteins
CC and/or HmW peptides. The nucleotide sequences encoding the HMW proteins
CC can be used to isolate and clone hmw genes from other non-typeable
CC strains of Haemophilus via hybridisation reactions. The present sequence
CC represents an HMWA protein from a non-typeable strain of H. influenzae.
XX (Updated on 12-SEP-2003 to standardise OS field)

XX
SQ Sequence 1011 AA;
Query Match 5.9%; Score 111.5; DB 3; Length 1011;
Best Local Similarity 20.3%; Pred. No. 0.63;
Matches 68; Conservative 65; Mismatches 147; Indels 55; Gaps 14;

QY 60 ESHTYDRMTFLCLSHNTINGACPTSPSSSVSGETNITLQFTKRS-LIKRELQIK 118
DB 363 EKNAIFSTHNLITLGGNVTLGGSSNKGNNINSKANVTLQAEAGTSLDKKE---- 424
QY 119 GYKQLLFKSNCPBGLTL--NSAHFNCNKNAASGASLYLIPAGEL-KNLPFGGIW--DA 173
DB 425 --RTLTLGNVSVGGNLIIGSNNAHIDGNLS-ABSAPK-----QCKTNNNLNITGFTNNG 477
QY 174 TLKLRVRRYSEYTYGTYTINITIKLTD-----KGNL-----QIWLDPFKSDARVD 218
DB 478 TADINIKQGVVKLQGDITNNGNLTNNITNASVNQKTIINGNITNKKGLNLIKDIKANAEIQ 537
QY 219 L--NLPTGGGTIGRNSVDMCFYDGYSTNSSLERIQDNNPKSD-GKTYLRKINDDTK 275
DB 538 IGGNISQKGNLTISDDKINI-----TKRIEKADTDQGNSDSGVASNANLTIKTK 588
QY 276 EIAYTSLLLAG---KSLTPTNGTSLNIADAASLETNWSITAVTMEPEISVPVLCWPGRL 332
DB 539 ELTLDNLNIGSFNKAIBITAKDNDLILGKASSDNSNAKQITFDKVKDKSIS--AGNHNV 646
QY 333 QLDKAVENPEA-----GOYMGNNINVTFTPSQTL 361
DB 647 TLNSKVETSDSGTNGSDDDNIGLITISAKDVTV 681

RESULT 14
AAB01832
ID AAB01832 standard; protein; 1529 AA.
XX AAR41732;
AC AAR41732;
XX 25-MAR-2003 (revised)
DT 26-APR-1994 (first entry)
XX High molecular weight protein 4 (HMW4).
DE

CC and/or HmW peptides. The nucleotide sequences encoding the HmW proteins
CC can be used to isolate and clone hmw genes from other non-typeable
CC strains of Haemophilus via hybridisation reactions. The present sequence
CC represents a mature HmW protein from a non-typeable strain of H.
CC influenzae. (Updated on 12-SEP-2003 to standardise OS field)

XX
SQ Sequence 1005 AA;
Query Match 5.9%; Score 111.5; DB 3; Length 1005;
Best Local Similarity 20.3%; Pred. No. 0.62;
Matches 68; Conservative 65; Mismatches 147; Indels 55; Gaps 14;

QY 60 ESHTYDRMTFLCLSHNTINGACPTSPSSSVSGETNITLQFTKRS-LIKRELQIK 118
DB 363 EKNAIFSTHNLITLGGNVTLGGSSNKGNNINSKANVTLQAEAGTSLDKKE---- 418
QY 119 GYKQLLFKSNCPBGLTL--NSAHFNCNKNAASGASLYLIPAGEL-KNLPFGGIW--DA 173
DB 425 --RTLTLGNVSVGGNLIIGSNNAHIDGNLS-ABSAPK-----QCKTNNNLNITGFTNNG 471
QY 174 TLKLRVRRYSEYTYGTYTINITIKLTD-----KGNL-----QIWLDPFKSDARVD 218
DB 472 TADINIKQGVVKLQGDITNNGNLTNNITNASVNQKTIINGNITNKKGLNLIKDIKANAEIQ 531
QY 219 L--NLPTGGGTIGRNSVDMCFYDGYSTNSSLERIQDNNPKSD-GKTYLRKINDDTK 275
DB 532 IGGNISQKGNLTISDDKINI-----TKRIEKADTDQGNSDSGVASNANLTIKTK 582
QY 276 EIAYTSLLLAG---KSLTPTNGTSLNIADAASLETNWSITAVTMEPEISVPVLCWPGRL 332
DB 583 ELTLDNLNIGSFNKAIBITAKDNDLILGKASSDNSNAKQITFDKVKDKSIS--AGNHNV 640
QY 333 QLDKAVENPEA-----GOYMGNNINVTFTPSQTL 361
DB 641 TLNSKVETSDSGTNGSDDDNIGLITISAKDVTV 675

RESULT 13
AAB01832
ID AAB01832 standard; protein; 1011 AA.
XX AAB01832;
XX 12-SEP-2003 (revised)
DT 11-SEP-2000 (first entry)
XX Haemophilus influenzae strain K21 HMW2A protein, SEQ ID NO:39.
XX HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
XX non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
XX recombinant production; Escherichia coli; antibacterial; vaccine;
XX human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
XX detection; diagnosis.
XX Haemophilus influenzae; strain K21.
XX WO200020609-A2.
XX 13-APR-2000.
XX 07-OCT-1999; 99WO-CA000938.
XX 07-OCT-1998; 98US-00167568.
XX 08-DEC-1998; 98US-00206942.
XX (CONN-) CONNAUGET LAB LTD.
XX Loosmore SM, Yang Y, Klein ME;
XX WPI; 2000-303789/26.
XX N-PSDB; AAA52181.
XX Nucleic acid molecule for producing recombinant high molecular weight

Qy 307 ETNKNRITAVTPEISVZVLCPGP-RLQLDAKVENPEAQYMGINVT 353
 : : || : : : : : : : : : :
Db 1074 NADAKK--VTFDKVKDSKSTDGHNVTLNSEVKTSCGSSNAGNDNST 1119

Search completed: May 6, 2004, 10:17:42
Job time : 61 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

QM protein - protein search, using sw model

Run on: May 6, 2004, 10:15:31 ; Search time 20 Seconds
(without alignments)
1736.257 Million cell updates/sec

Title: US-09-839-894-10

Perfect score: 1886
Sequence: 1 MNKILFIFTLFFSVLTFFA.....EAGQYMGNIWTFPPSSQTL 361

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76.*
1: p.r1.*
2: p.r2.*
3: p.r3.*
4: p.r4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1558.5	82.6	360	2	D56617	cfag protein precu
2	949.5	50.3	363	2	S49539	CotD protein precu
3	930.5	49.3	364	2	S57937	CotD protein precu
4	275.5	14.6	359	2	A80541	probable fimbrial
5	120	6.4	1335	2	G90975	probable factor fi
6	120	6.4	2660	2	E85822	probable invasin Z
7	114	6.0	650	2	S28835	alpha-agglutinin -
8	109	5.8	2383	2	D4962	probable membrane
9	108	5.7	524	2	S55097	probable membrane
10	108	5.7	2529	2	B64635	toxin-like outer m
11	107	5.7	304	2	A64904	probable fimbrial
12	107	5.7	304	2	C90892	probable adhesin (
13	107	5.7	304	2	F85725	probable adhesin,
14	106	5.6	671	2	A38109	autolysin - Entero
15	105.5	5.6	843	2	AC2507	hypothetical prote
16	105	5.6	598	2	AB1236	internalin protein
17	104.5	5.5	1441	2	B86807	hypothetical prote
18	103.5	5.5	691	2	B75622	hypothetical prote
19	103.5	5.5	1269	2	A90267	proteinase related
20	103.5	5.5	2399	2	H71879	toxin-like outer m
21	102	5.4	692	2	G90284	hypothetical prote
22	101	5.4	1238	2	AH0038	probable exported
23	100	5.3	1752	2	T48965	hypothetical prote
24	99.5	5.3	5627	2	C83339	hypothetical prote
25	99	5.2	802	2	AH1580	phenylalanyl-tRNA
26	99	5.2	856	2	B81399	probable periplasm
27	99	5.2	1849	2	C41859	Iga-specific metal
28	98.5	5.2	608	2	H64473	hypothetical prote
29	98.5	5.2	4919	2	T31105	hypothetical prote

30 98 5.2 587 2 AC1510 internalin protein
31 98 5.2 758 2 S47866 Spt21 protein - ye
32 98 5.2 1014 2 S37405 cytotoxic necrotiz
33 98 5.2 1157 2 S38160 NUP-33 protein - y
34 97.5 5.2 1797 2 F69195 cell surface glyco
35 97 5.1 454 2 T01337 hypothetical prote
36 96.5 5.1 2364 2 I40884 cytotexin L - Clo
37 96 5.1 992 2 T28421 probable DNA-direc
38 95.5 5.1 713 1 ALB8XR cyclomaltodextrin
39 95.5 5.1 895 2 T02597 Mutator-like trans
40 95 5.0 413 1 S28602 translation releas
41 95 5.0 561 2 A84113 transposase (12) B
42 95 5.0 2334 2 S32920 cell wall-associat
43 95 5.0 5188 2 B85547 probable RTX fami
44 94.5 5.0 835 2 E7-691 outer membrane ass
45 94.5 5.0 853 2 T47223 replication licens

ALIGNMENTS

RESULT 1

D56617
cfag protein precursor - Escherichia coli plasmid NTP113
C:Species: Escherichia coli
C:Date: 05-Jan-1996 #sequence_revision: 05-Jan-1996 #text_change: 10-Dec-1999
C:Accession: D56617
R:Jordi, B.J.; Willshaw, G.A.; van der Zeijst, B.A.; Gastra, W.
DNA Seq. 2, 257-263, 1992
A:Title: The complete nucleotide sequence of region 1 of the CFA/I fimbrial operon of hum
A:Reference number: A56617; MUID:9232981; PMID:1352712
A:Accession: D56617
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-360 <J03>
A:Cross-references: GB:M55661; NID:G145507; PIDN:AAC41417.1; PID:G145511
A:Experimental source: enterotoxigenic strain, CFA/I-SI plasmid NTP113
A:Note: sequence extracted from NCBI backbone (NCB-N:138960, NCBIP:108972)
C:Genetics:
A:Gene: cfag
A:Genome: plasmid
C:Superfamily: Escherichia colonizing factor antigen cfagE

Query Match	82.6%	Score 1558.5	DB 2	Length 360
Best Local Similarity	81.2%	Pred. No. 1.4e-109		
Matches 293	Conservative 31	Mismatches 36	Indels 1	Gaps 1
Qy	1	MNKLIFITLFFSVLTFEAYSADKIPCESITNLIFGPRDRNESSPKNIINNHITAYSE	60	
Db	1	MNKLIFITLFFSSGFFTFAYADKXNPGSENNTNTIGPHDRGGSSPIYNILNSYUTAYNG	60	
Qy	61	SHTLYDRMTFCLSSHNITNGACPTSENPSSSVSGSETNITLQFTKESLTKRELIQKY	120	
Db	51	SHLYDRMSFCLSSQNTINGACPSDDAPGTATIDGETNITLQFTKESLTKRELIQKY	120	
Qy	121	KQLLPKSVNCSGLTNSAHFNCKNKAASGLSYIYPAGELKNLPQGIWDATLKLAVK	180	
Db	121	KQFLFNANCPSKALNSHFQCNREQASGATLSLYIPAGELKNLPQGVVNAVLNVK	180	
Qy	181	RRSYETGYTINITIKLTDKGNTQIWLPOKSDARVDNLNRPPTGGTYIGNSVDMCFY	240	
Db	181	RRYDTYTYTINITINVTLDKNTQIWLPOKSNARVDNLNRPPTGGTYIGNSVDMCFY	240	
Qy	241	DGYSTNSSLEIRFODNNPKSDGKPYLKKINDTKEIAYTLLSLAGKSLPTNGTSLNI	300	
Db	241	DGYSTNSSLEIRFODDNNKSDGKPYLKKINDDSKELIVTLLSLAGKSLPTNGQALNI	300	
Qy	301	ADAASLETNWKRIITAVTTPETISVPVLCWPGRLQDAKVENPEAGQYMGNIWTFPPSSQT	360	
Db	301	-NTASLETNWNRIITAVTTPETISVPVLCWPGRLQDAKVENPEAGQYMGNIKITFTPPSSQT	359	
Qy	361	L 361		

Db 360 L 360

RESULT 2

S49539

C: protein precursor - Escherichia coli

C: Species: Escherichia coli

C: Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 10-Dec-1999

C: Accession: S49539

R: Froehlich, B.J.; Karakashian, A.; Melsen, L.R.; Wakefield, J.C.; Scott, J.R.

Mol. Microbiol. 12, 387-401, 1994

A: Title: CooC and CooD are required for assembly of CS1 pili.

A: Reference number: S49539; PMID:94344028; PMID:7915003

A: Accession: S49539

A: Status: preliminary

A: Molecule type: DNA

A: Residues: 1-363 <PRO>

A: Cross-references: EMBL:X76908; NID:G488735; PIDN:CAA54230.1; PID:G488737

C: Superfamily: Escherichia colonizing factor antigen cfaE

Query Match

Best Local Similarity 50.3%; Score 949.5; DB 2; Length 363;

Matches 197; Conservative 55; Mismatches 98; Indels 21; Gaps 11;

QY 3 KILFIPTLFFSVLFTFAVSADKIPGDE--STNIF-GER-DRNBSSPKHNLNNHITAY 58

Db 2 KKIFIFL----SIITSAVVSAGRYPETTVGNLTKGFQAPRLDRSQSPTYNIFTNHVAGY 57

QY 59 SESHTLYDRMTFLCLSSHTNLGACPTSENPSSSVS--GETNITLQTEKRSLLIKRELOJ 117

Db 58 SLSHSLYDRIVFLCTSSSPVNGACPTI---GTSGVQYQYTTTTITLQTEKRSLLIKRNKL 114

QY 118 KGVKOLLPKSVNCPGS--LTLASAHFNCKNA--ASGASLYLIPAGELKNLPFGGIWDAT 174

Db 115 AGNKPIWENQCDRSNLVMSKSWSCAHNGAKTLLNLIYPAGEINKLPFGGIWEAT 174

QY 175 IKLRVKRRYSEY---YGTITNITIKLTDKGNIQIWLPOKSDARVDNLNRPFGGTVI 230

Db 175 LILRLS-RYGEVSSYHYGNYTNITVDLTDKGNIQIWLPGPHSNPRVDNLNRPIGNKYK 233

QY 231 GNSVDMCFYDGYSTNSLSLEIRFQDNNPKSDGKFYLRKINDDTKEIATLSLLAGKSL 290

Db 234 GNSLDMCFYDGYSTNSDSMVVKFQDNNPTNSSEVNLKIG-GTEKLPYAVSLMGEKIF 292

QY 291 TPNGTSLNIAAASLETWNRTITAVTPEISVPVLCWPGRIQLDAKVENPAGQVMGNI 350

Db 293 YPNGQSFTINDSSVLETNWNRVAVAMEVNVVPCWPARILLNADVNPAGQVSGOI 352

QY 352 NVITFPSSQTL 361

Db 353 YITFPSEVNL 363

RESULT 3

S57937

C: protein precursor - Escherichia coli

C: Species: Escherichia coli

C: Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 10-Dec-1999

C: Accession: S57937

R: Froehlich, B.J.; Karakashian, A.; Melsen, L.R.; Scott, J.R.

submitted to the EMBL Data Library, January 1995

A: Description: The genes for CS2 pili of enterotoxigenic Escherichia coli and their inte

A: Reference number: S57934

A: Accession: S57937

A: Status: preliminary

A: Molecule type: DNA

A: Residues: 1-364 <PRO>

A: Cross-references: EMBL:Z47800; NID:G897725; PIDN:CAA87763.1; PID:G897729

C: Genetics:

A: Start codon: TGG

C: Superfamily: Escherichia colonizing factor antigen cfaE

Query Match

49.3%; Score 930.5; DB 2; Length 364;

Best Local Similarity 48.4%; Pred. No. 2.1e-62;

Matches 177; Conservative 64; Mismatches 118; Indels 7; Gaps 3;

QY 1

Db 1 MKKILFIPTLFFSVLFTFAVSADKIPGDESIITNIFGPRDRN--ESSPKHNLNNHITAY 58

Db 1 MKKIVFLSKFLCQVYQSWHTNVEAGSINKTSIGPIDRSAAASYFAHYIFHEHVAGY 60

QY 59 SESHTLYDRMTFLCLSSHTNLGACPTSENPSSSVSGETNITLQTEKRSLLIKRELOIK 118

Db 61 NKDHSLEPRMTFLCWSSTDAKSGACPTGSKSS--QGETNKLIFTEKSKLARKTLNLK 118

QY 119 GKQLLFKSVNC---PSGLTNSAHFNCKNAASGASLYLIPAGELKNLPFGGIWDATL 175

Db 119 GYKGFLEYSDRCIHYVDKWNLSHTVKCVGSGFTGVDFTLXIPQGEIDGLTGGWEATL 178

QY 176 KLRVKRRYSEYGTITNITIKLTDKGNIQIWLPOKSDARVDNLNRPFGGTVIGRSV 235

Db 179 ELRVKRVHYDHYGNYKVNITVDLTDKGNIQWTFKFSDDPRIDLNLREGENKYSNVL 238

QY 236 DMCFYDGYSTNSLSLEIRFQDNNPKSDGKFYLRKINDDTKEIATLSLLAGKSLTPNG 295

Db 239 EMCLYDGYSTHSQSIEMRFQDSDQTNNEYNLIKTEBPLKXLPYKLSLLLGREFYPNG 298

QY 296 TSLNIADAASLETWNRTITAVTPEISVPVLCWPGRIQLDAKVENPAGQVMGNI 355

Db 299 EAFITNDTSSLFINWNRKSVSLPQISIPVLCWPAFLTFMSELNPNPEAGEYSGLNVTFT 358

QY 356 PSSQTL 361

Db 359 PSSSL 364

RESULT 4

A:0541

probable fimbrial protein tocd [imported] - Salmonella enterica subsp. enterica serovar 7

C: Species: Salmonella enterica subsp. enterica serovar Typhi

A: Note: this species has also been called Salmonella typhi

C: Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C: Accession: AE0541

R: Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A: Authors: Park, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;

A: Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A: Reference number: AB0502; PMID:21534947; PMID:11677608

A: Accession: AE0541

A: Status: preliminary

A: Molecule type: DNA

A: Residues: 1-359 <PAR>

A: Cross-references: GB:AL513382; PIDN:CAD08773.1; PID:G16501589; GSPDB:GN00176

C: Genetics:

A: Gene: tocd

Query Match

14.6%; Score 275.5; DB 2; Length 359;

Best Local Similarity 29.7%; Pred. No. 3.2e-13;

Matches 90; Conservative 48; Mismatches 130; Indels 35; Gaps 11;

QY 70 FLCLSSHTNLGACPTSENPSSSVSGETNITLQTEKRSLLIKRELOIKGYKQLFKSVN 129

Db 79 WVCNRNENEGACETHLVWVYAFGAVSKTLRFRQISHAEITL-----ILLGSVR 131

QY 130 --CPSGJTLNSAHFNCKNAAS--GASLYLIPAGELKNLPFGGIWDATLKL-RVKRY 183

Db 132 DACTYTG-----INWNAACQWGRSLKLRIPSEELAKIPTSGTWKATLVLDLWMSG 183

QY 184 STYGTITNITIKLTD--KGNIQIWLPOK--KSDARVDNLNRPFGGTVIGRSVDMCY 240

Db 184 DDPGLTSTDTITLVNTHFAENALYFPFGFATPRVDNLNHRNANASQMSGRANDMCLY 243

QY 241 DGSYNSLSLEIRFQDNNPKSDGKFYLRKINDDTKEIATLSLLAGKSLTPNGTSLNI 300

Db 244 DG-GVKARSLQMKLEGSNKSSTGFGQVIXSDSADT--IDYAVSMNYGGRSIPVTRGVFSL 300

A:Molecule type: DNA
A:Residues: 1-650 <DEH>
A:Cross-references: EMBL:X87611; NID:G854567; PIDN:CAA60926.1; PID:G854577
Ride Haan, M.; Grievell, L.A.; Smits, P.H.M.
submitted to the Protein Sequence Database, September 1995
A:Reference number: S56771
A:Accession: S57019
A:Molecule type: DNA
A:Residues: 1-650 <ZAG>
A:Cross-references: EMBL:Z49504; NID:G1015625; PIDN:CAA89526.1; PID:G1015626; MIPS:YJR00
C:Genetics:
A:Gene: SGD:SAG1; AGALL1; AGALPHAI
A:Cross-references: SGD:S0033764; MIPS:YJR004c
A:Map position: 10R
C:Keywords: glycoprotein

Query Match 6.0%; Score 114; DB 2; Length 650;
Best Local Similarity 20.9%; Pred. No. 0.95;
Matches 95; Conservative 52; Mismatches 147; Indels 172; Gaps 23;

QY 4 ILFIPLTFSSVL-----FTFA-----VSADKIPG-----DESITNIFGRDRNE--- 43
DB 8 ILWLFSLALASAININDITFSNLEITPLTANKQPDQGWATATDFSIADASSIREGDEFTL 67
QY 44 SSP-----KKNILNNHITA-----YSESHLYDRMTFLC-----LSSHNTLN 80
DB 68 SMFHVYRIKLLNSSQAT-SLADGTEAFKQYVSQQAAYLYENTFTCTAQNDLSSYNTID 127
QY 81 GACPTSENSSSVSGETNITLQFTKRSILKRELOIKGYQLFKSVNC-PSGLTLNSA 139
DB 128 GSITFLNFSGGSSVEYL-----ENAKFKSGPMVLKGMQSDVDFDAAPTENVF 182
QY 140 HFNCNXXAASGASLYL--YIPAGELKNLPFGGIWDATLKLVRKRRYSETYGTVTINITIK 197
DB 183 HSGRSTGYSGFSYHLGMYCNGYF-----LGG-----TEKI-----DYDSSNNVDL- 225
QY 198 LIDKGNIOI-----WLQPKSDARVDLNLPRPGGTYIGRN---SVDMCFYDG--- 242
DB 226 --DCSSVQVYSSNDPNDMFPQSYNDTADV-----TCFGSNLWITLDEKLYDGEWL 275
QY 243 -----YSTNSSSLE--IRPDNN--PKSDGKF 265
DB 276 WYNALQSLPANVNIIDHAEFFQVTCIDTANTTYATQFSTRFIVYQGRNLGTAASKS 335
QY 266 YLRKINDYKEI---AYTL-----LLAGKSLTPNGTSLNTADAA 304
DB 336 FISTTTDLTISNTAISAYTSGISTVETGNRTTSEVISHVVTGKLSPTATTSLT-AQTS 395
QY 305 SLETNV-----RITATMPEISVPVLCWQGR 332
DB 396 IYSTDSNITVGTDLHTTSEVISDVETISRETASTVVAAPTSTTGWTCAM 444

RESULT 8
D64962
probable membrane protein b1978 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: D64962
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: D64962
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2383 <BLAT>
A:Cross-references: GB:AE000289; GB:U00096; NID:G1788285; PIDN:AACT5042.1; PID:G1788289;
A:Experimental source: strain K-12, substrain MG1655
C:Keywords: nucleotide binding; P-loop; transmembrane protein
F:54-70/Domain: transmembrane #status predicted <TM>
F:1564-1571/Region: nucleotide-binding motif A (P-loop)

Query Match 5.8%; Score 109; DB 2; Length 2383;
Best Local Similarity 20.0%; Pred. No. 12;
Matches 78; Conservative 65; Mismatches 157; Indels 90; Gaps 15;

QY 13 SSVLTFFAVSADKIPQDESITNIFGRDRNESSPXNITLNHITAY---SESHLYDRMT 69
DB 762 SAKIATLSASNGVLANNAANTVSVNVADEGS---NPINDHTVTFVLSGATSENNON 818
QY 70 FCLSSHNTLNGACPTSENSSSVSGETNITLQFTKRSLI-----K 112
DB 819 ----TAKTDVNGLA-TFDLKSSKQEDNTVEVILENGVKOTLIVSFVGDSSSTAQVDLQSK 873
QY 113 R3LQIKGYKQL-LFKSYNCPGSLTNSAHFNCKNAASGASILYIPAGELKNLPFGGIW 171
DB 874 NEVVADGNDSVTMTATVDAKGNLLNDVMVTNVNVAZAKLSQTEVNSHD-----GIA 926
QY 172 DATLKLVRKRRYSETYGTVTINITIKLTDKGNIQWLPOFKSDARVDLMLRPTGGTYIG 231
DB 927 TAILT-----SLKNGDYRVTAHSVSSGQANQVNFQDSTAALTSLV-PSGDTIV-- 976
QY 232 RNSVDVCFYDGYSTKSS-----SLRIFQDNPNKSPGKFLYLRKINDDKETIATLSLLAG 287
DB 977 -----TNTAPQYMTATLQDKNGNPLKDKETFSVPND---VASKFSTSNCG 1019
QY 289 KSLTPTNGTSL-----NIADAASLETNWRNLTAYTMPEISVPVLC 327
DB 1020 KGMTDSNGVALASLTGTLTLAGTHIMARLANSVSDAQMPTFVADKORAVVVLQTSKAEII 1079
QY 328 WFG--RLQLDARVENPEAGQYMGINVTPT 355
DB 1080 GNGVDETILTATVKOP-SNHPVAGITVNET 1108

RESULT 9
S55097
probable membrane protein YMR215w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YMR261.09
C:Species: Saccharomyces cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
C:Accession: S55097
R:Dedman, K.; Brown, D.; Bowman, S.
submitted to the EMBL Data Library, June 1995
A:Reference number: S55097
A:Accession: S55097
A:Molecule type: DNA
A:Residues: 1-524 <DED>
A:Cross-references: EMBL:Z49809; NID:G854459; PID:G854467; GSPDE:GN00013; MIPS:YMR215w
A:Experimental source: strain AB972
C:Genetics:
A:Gene: MIPS:YMR215w
A:Cross-references: SGD:S0004828
A:Map position: 13R
C:Superfamily: glycopospholipid-anchored surface glycoprotein GASI
C:Keywords: transmembrane protein
F:7-23/Domain: transmembrane #status predicted <TM>

Query Match 5.7%; Score 108; DB 2; Length 524;
Best Local Similarity 24.7%; Pred. No. 2;
Matches 69; Conservative 42; Mismatches 96; Indels 72; Gaps 18;

QY 119 GYKQL--LFKSYNCPGSLTNSAHFNCKNAA---SGASILYIPAGELKNLPFGGIWDA 173
DB 264 GYDKLNSTFEDAVIP---LIFSEYGCNKNTRPTDFEVESEGLY---GGLNVPFSGSL--- 313
QY 174 TLKLVRKRRYSETYGTVTINITIKLTDKGNIQWLPOFKSD-ARVDLNL--PTGGGT 228
DB 314 -----VVEYTEANNYGL---VKLDDSGSL-----TYKDDFVNLSQLKNVLPFTKES 359
QY 229 YIGRNVDMVC-----FYDGYSTNSSSLE-----IRFQDNPKSDGK----- 264
DB 360 EISSDSIYKCDNSAINTNYSVGFTNNFTLPSQPAETANNMIEYGVNGTNT-GKILTDYAVP 418

```
QY 265 ----PYLRKINDTKEIATYLSLLLAGKSLTPTNGFSLNIAADAASLETNWR-ITAVTMP 319
DB 419 TTFNYTIKNNKDDT--ISATISYDKA-NSLNLDTTATTVAASASTOSSSSSLTSSP 475
QY 320 ETSVPVLCWFGRLQLDKVENPEAGYMGNI-NVFTPTPS 357
DB 476 SSTGSSSTG-----SSSASSSSKSGVGNIVNVFSQS 510

RESULT 10
B64635
toxin-like outer membrane protein HP0922 - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C>Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C:Accession: B64635
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKen-
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Praser, C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467; PMID:9252185
A:Accession: B64635
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2529 <TOM>
A:Cross-references: GB:AR000602; GB:AE000511; NID:G2314060; PIDN:AD07969.1; PID:G231406

Query Match 5.7%; Score 108; DB 2; Length 2529;
Best Local Similarity 22.5%; Pred. No. 16;
Matches 84; Conservative 50; Mismatches 122; Indels 118; Gaps 22;

QY 85 TSENTESSSVSGETNITLQFTKRSLL-----KRELQIKYKQLLFK--SVNCP 131
DB 977 TSNFNATTQLGNTNFTL---SSQSLNFNGDPTTLQNNANITLGNKSQAFFKNSLTLDNN 1033

QY 132 SGTLL-NSAFENCKNAA--SGASLYLYT-PA-GELKNLPFGGIWDATLRLVKRRYSEY 187
DB 1034 SNLSLDNQSVLNANTSAFNKQASLNYNGSOATFNSLFPNG---GTLSELNASSKLNAGN 1090

QY 188 GYVITNITIKLTD-----KGNQIWLFPQKSDARVDLNLRLPTGGGYVIGRNSVD 236
DB 1091 AFSFNNTNINLDDSVLSASNTSLNANI-----NFGASQADF-----GNTIITDASEN 1140

QY 237 MCFYGYVSNSSSL-----BIRFOONPK-----SDGKFYLRKIN-----DDT 274
DB 1141 --FDSASSLNFNNTANGALNFNGYTPSLTKALMSVSGQFVLNNGDINLSINIFDNI 1198

QY 275 KEIATYLSLLAGKSLTPTNGTS-----LNIAAASLETN---WNRTAVTMP 321
DB 1199 KSVTY--NIINAKQGITGSGANGYKILFYGMKIQNATYSDNNITQWSFNPINSSQI 1256

QY 322 -----SVPVL-----CWFGRLQLDKVENP-----EAGQY--M 347
DB 1257 IQBSIKNGDLTIEVLNPNKSNASNTIENIAPELNYQASKNQNGPYSDYSDNQAGTYILT 1316

QY 348 GNINVTFTP-SSOT 360
DB 1317 SNIKGLFTPKGSOT 1330

RESULT 11
A64904
probable fimbrial protein b1502 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: A64904
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, Y.; Co-
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: A64904
```

```
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-304 <BLAT>
A:Cross-references: GB:AE000247; GB:U00096; NID:G1787773; PIDN:AACT4575.1; PID:G1787779;
A:Experimental source: strain K-12, substrain MG1655
C:Superfamily: fimbrial protein fimH
C:Keywords: fimbria
```

```
Query Match 5.7%; Score 107; DB 2; Length 304;
Best Local Similarity 21.0%; Pred. No. 1.2;
Matches 75; Conservative 51; Mismatches 125; Indels 106; Gaps 19;
```

```
QY 3 KILF-IFTLFPSSVLFTFAVSADKIPGDESI---TNIFGPRDRNESSPKHNI---LNNH 54
DB 3 KVLFGIYLLLMAGKVAFSCNVD---GGSSICAGTTSVYVNL-D-PVIQGNLVVDLSOH 63
QY 55 IPAYSESHLYDRMTFLCLSSH-NTLNGACPTSENPSSSVSGETNITLQFTKRSLLIK 113
DB 64 ISCWNDYGGWYD-----TDHINLVQ-----SAFAG----- 89
QY 114 ELQIKGYK-QLLFKSVNCPSEGLTNSAHFNCKNKAASGASLYLYIPAGELKNLPFGGIWD 172
DB 90 --SLQSYKGSLYMNVVTFPPLTTNTNVLIDIGKTPMPLPLKLYI-----TPVGAAG 140
QY 173 ATLK-----LRVKRRYSEYGT-----YTINITLK-----LTDKGNIQIWLPPCK 212
DB 141 VIKAGEVIARIHMTKIALGSGNPNFTWNIISNNVMPGCGTYDSENVTVDLPDP 200
QY 213 SPARVDLNLRLPTGGGYTIGRNSVDMCFYGYSTNSSSLEIRFQDNNPKSDGKFYLRKIND 272
DB 201 GSAETPL-----GVYCS-SBQKLSFY--SGATTDSSROV-FANTAP-----D 239
QY 273 DKETAATLSLLLAGK-----SLTPNGISLNIADAASLETNWNRITAVTMP 321
DB 240 ATKASGVGTLMRNGKILATGENVSLGTVYKSKVPLGLSATYGTGKNKVSAGEVQSV 296
```

RESULT 12

```
C90892
probable adhesin [imported] - Escherichia coli (strain O157:H7, substrain RMD 0509952)
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: C90892
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: C90892
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-304 <RAY>
A:Cross-references: GB:BA000007; PIDN:BA835530.1; PID:G13361573; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: ECs2107
C:Superfamily: fimbrial protein fimH
```

```
Query Match 5.7%; Score 107; DB 2; Length 304;
Best Local Similarity 21.0%; Pred. No. 1.2;
Matches 75; Conservative 51; Mismatches 125; Indels 106; Gaps 19;
```

```
QY 3 KILF-IFTLFPSSVLFTFAVSADKIPGDESI---TNIFGPRDRNESSPKHNI---LNNH 54
DB 8 KVLFGIYLLLMAGKVAFSCNVD---GGSSICAGTTSVYVNL-D-PVIQGNLVVDLSOH 63
QY 55 IPAYSESHLYDRMTFLCLSSH-NTLNGACPTSENPSSSVSGETNITLQFTKRSLLIK 113
DB 64 ISCWNDYGGWYD-----TDHINLVQ-----SAFAG----- 89
QY 114 ELQIKGYK-QLLFKSVNCPSEGLTNSAHFNCKNKAASGASLYLYIPAGELKNLPFGGIWD 172
DB 90 --SLQSYKGSLYMNVVTFPPLTTNTNVLIDIGKTPMPLPLKLYI-----TPVGAAG 140
```

173 ATLK-----LRVKRYSETYGT-----YTINITIK-----LTDKGNQIWLPOPK 212
Db 141 VVIKAGEVIARIMYKIALGSGNPRFTWNIISNNSVMTPTGGCTVDSRNTVNLDPDP 200
Qy 213 SDARVDNLNRPRTGGGYPIGRNSVDMCFYDGYSTNGSSLEIRFQDNNPKSDGKFLYLRKIND 272
Db 201 GSAEIP-----GVYCS-SEQLSFLSGTTDSARQV-FANTAP-----D 239
Qy 273 DTKEIAYTLLSLLAGK-----SLTPNGTSLNIADAASLETNNR-TAVTMPEI 321
Db 240 ATKAGVGVSIMRNGKILATGENVSLGTWNKSKVPLGLSATYGTQGNKVSAGTVQSV 296

RESULT 13
F85725
probable adhesin, Fim4 type protein Z2206 [imported] - Escherichia coli (strain O157:H7,
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: F85725
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Porarousis, K.; Apodaca,
Nature 405, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; PMID:21074935; PMID:11206551
A:Accession: F85725
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-304 <SPO>
A:Cross-references: GB:AE005174; NID:gl2515169; PIDN:AAG56266.1; GSPDB:GN00145; UWGP:Z22
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z2206
C:Superfamily: fimbrial protein fimH

Query Match 5.6%; Score 107; DB 2; Length 304;
Best Local Similarity 21.0%; Pred. No. 1.2;
Matches 75; Conservative 51; Mismatches 125; Indels 106; Gaps 19;
Qy 3 KILP-ITLTFSSVLFTFAVSADKIPQDESI---TNIFGRDRNESSPKHNI---LNNH 54
Db 8 KVLFGIYLLLMAGKVFAPSKVD---GSSIGACTTSVYVNLDPVITQPNQLVVDLSQH 63
Qy 55 ITAYSESHLYDRMTFLCLSSH-NTLNAGCAPTSENPSVSSVSGETNITLOFTEKRSILIKR 113
Db 64 ISCNWDYGGWYD-----TDHINLVG-----SAPAG----- 89
Qy 114 ELQIKGYK-QLLFKSVNCPGSLTNSAHFNCKNAAGASLYLYIPAGELKNLPFGGIWD 172
Db 90 -S-OSYKGSGLYNNVTPPLTNTNVLDTGDKTPEPLKLYI-----TPVGAAGG 140
Qy 173 ATLK-----LRVKRYSETYGT-----YTINITIK-----LTDKGNQIWLPOPK 212
Db 141 VVIKAGEVIARIMYKIALGSGNPRFTWNIISNNSVMTPTGGCTVDSRNTVNLDPDP 200
Qy 213 SDARVDNLNRPRTGGGYPIGRNSVDMCFYDGYSTNGSSLEIRFQDNNPKSDGKFLYLRKIND 272
Db 201 GSAEIP-----GVYCS-SEQLSFLSGTTDSARQV-FANTAP-----D 239
Qy 273 DTKEIAYTLLSLLAGK-----SLTPNGTSLNIADAASLETNNR-TAVTMPEI 321
Db 240 ATKAGVGVSIMRNGKILATGENVSLGTWNKSKVPLGLSATYGTQGNKVSAGTVQSV 296

RESULT 14
A38109
autolysin - Enterococcus faecalis
C:Species: Enterococcus faecalis
C:Date: 24-Jul-1992 #sequence_revision 24-Jul-1992 #text_change 15-Oct-1999
C:Accession: A38109
R:Beliveau, C.; Potvin, C.; Trudel, J.; Asselin, A.; Bellemare, G.
J. Bacteriol. 173, 5619-5623, 1991
A:Title: Cloning, sequencing, and expression in Escherichia coli of a Streptococcus faec

A:Reference number: A38109; PMID:91358349; PMID:1679432
A:Accession: A38109
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-671 <BEL>
A:Cross-references: GB:M58002; NID:gl53658; PIDN:AAA67325.1; PID:9829194

Query Match 5.6%; Score 106; DB 2; Length 671;
Best Local Similarity 23.6%; Pred. No. 3.9;
Matches 73; Conservative 39; Mismatches 103; Indels 94; Gaps 17;
Qy 41 RNESPKHNI-LNNHITAYSESHLYDRMTFLCLSSHNTLNGCAPTSENPSVSSV----- 94
Db 309 RYATDPSYNAKLNNTAY--NLTYDTPSSGNTGGGTWNFGTGGSNKQSGNTTYTVK 366
Qy 95 SGET--NITLQF-----TEKRSI--IKRELQIKGYKQLLFKSVNCPGSLTNSAHFNCKN 146
Db 367 SGETLNKTAAGYGVSVANLRSWNGISGDLIFVQKLIIVKKGK--SGNTGGSGGSGNN 423
Qy 147 AASGASLYLYIPAGELKNLPFGGIWDATLKLVRKRYSETYTYTINIITIKLTDKGNQI 206
Db 424 -QSGTNTTYTVKSGDTLN-----KIAAQYG-----VTV-----ANLRS 455
Qy 207 WLPOFKSDARVDNLN-----RPTGGGYTIGRNSVDMCFYDGYSTNGSSLSLIRFQDNN 258
Db 456 W-----NGISGDLIFVQKLIIVKKGTSNT-----GGSGNGS-----NNN 491
Qy 259 PKSDGKFLYLRKINDDTKEIA--YTIS-----LLAGKSLTPTNGTSLNIADA 303
Db 492 QSGTNTTYTIKSGDTLNKTAAGYGVSVANLRSWNGISGDLIFAGQKIIVKKGTSNTGGS 551
Qy 304 ASLETNNR 312
Db 552 SNGSGNNNQ 560

RESULT 15
AC2507
hypothetical protein all7235 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120ali
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AC2507
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; XUID:2595285; PMID:11759840
A:Accession: AC2507
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-843 <KUR>
A:Cross-references: GB:BA000020; PIDN:BAE78319.1; PID:gl7135773; GSPDB:GN00180
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all7235
A:Genome: plasmid

Query Match 5.6%; Score 105.5; DB 2; Length 843;
Best Local Similarity 18.9%; Pred. No. 5.8;
Matches 71; Conservative 67; Mismatches 127; Indels 111; Gaps 18;
Qy 77 NTLNAGCAPTSEN-----PSSSVSGENITLQ-----FT--EKRSILYREL 115
Db 297 NILDGLVFTQNHGKFGAVKIDAQSLNIQSSNLALSAIYTSNFGTPTGESIQLDYKV 356
Qy 116 QIKGYKQLLFKSVNCPGSL-TNSAHFN-----CNKNAASGASLYLYIP-----AG 160
Db 357 TIQGGQIATTTTAPSGLTINSNLKISGDTPSYANPDGLGGINITFSVSSGGGDIAG 416
Qy 161 ELKNLPPG--GIWD-----ATLKLVRKRYSETYTYTINIITIKLTDKGNQIWL 209
Db 417 KINNIIGLQGVFTVAGSGAGGNLPLENLNLIKGGASLGSSTIRSGQGNV----- 471

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 10:15:01 ; Search time 46 Seconds
(without alignments)
2476.132 Million cell updates/sec

Title: US-09-839-894-10
Perfect score: 1886
Sequence: 1 MNKILFFTLFFSSVLFTFA.....BAGQYMGNTVTFSSQTL 361

Scoring table: ELOSUM62
Gapcp 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREXBL.25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriopl.*
- 17: sp_archaea.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result- being printed, and is derived by analysis of the total score distribution.

SUMMARY35

Result No.	Score	Query Match	Length	DB ID	Description
1	1886	100.0	361	2 Q93G67	Q93G67 escherichia
2	950.5	50.4	363	2 Q848J5	Q848J5 escherichia
3	949.5	50.3	363	2 Q47115	Q47115 escherichia
4	930.5	49.3	364	2 Q47119	Q47119 escherichia
5	381	20.2	387	2 Q8KQJ7	Q8KQJ7 burkholderi
6	371	19.7	387	2 Q8KQJ0	Q8KQJ0 burkholderi
7	275.5	14.6	359	16 Q9XDS1	Q9XDS1 salmonella
8	272.5	14.4	245	16 Q83J01	Q83J01 shigella fl
9	256.5	13.6	210	16 Q7UBH8	Q7UBH8 shigella fl
10	127	6.7	1891	16 Q8RGK2	Q8RGK2 fusobacteri
11	115.5	6.1	700	2 Q8VU50	Q8VU50 chlamydia p
12	115.5	6.1	847	2 P71132	P71132 chlamydophi
13	114	6.0	865	15 Q8QJH7	Q8QJH7 human immun
14	113	6.0	2646	5 Q8I521	Q8I521 p-asmcdium
15	111.5	5.9	743	16 Q8XNW2	Q8XNW2 clostridium
16	111	5.9	2768	16 Q8E9G6	Q8E9G6 shewanella

17	110.5	5.9	589	16	Q3F994	Q8F994 leptospira
18	110.5	5.9	1052	17	Q8QDS0	Q8QDS0 methanoscara
19	110	5.8	265	2	Q8KGI5	Q8KGI5 rhizobium l
20	110	5.8	846	2	P71133	P71133 chlamydophi
21	109.5	5.8	2454	5	Q8EB02	Q8EB02 dictyosteli
22	109	5.8	807	5	Q8I7V5	Q8I7V5 drosophila
23	109	5.8	843	16	Q823X3	Q823X3 chlamydophi
24	108	5.7	872	2	Q9RLA0	Q9RLA0 rickettsia
25	108	5.7	942	16	Q823X1	Q823X1 chlamydophi
26	108	5.7	2529	16	Q25579	Q25579 helicobacte
27	107	5.7	304	16	Q8XAX2	Q8XAX2 escherichia
28	106.5	5.6	1012	16	Q8PEE7	Q8PEE7 xanthomonas
29	106.5	5.6	1065	16	Q8EZX0	Q8EZX0 leptospira
30	106	5.6	1937	5	Q8IE94	Q8IE94 plasmodium
31	105.5	5.6	430	3	Q9GVN2	Q9GVN2 blumeria gr
32	105.5	5.6	943	16	Q8YKQ8	Q8YKQ8 arabidopsis
33	105	5.6	598	16	Q8Y717	Q8Y717 listeria mo
34	105	5.6	868	16	Q823X4	Q823X4 chlamydophi
35	105	5.6	2931	5	Q86AE3	Q86AE3 dictyosteli
36	104.5	5.5	1441	16	Q9CPL1	Q9CPL1 lactococcus
37	104	5.5	304	16	Q8CW27	Q8CW27 escherichia
38	104	5.5	803	10	Q8AUS5	Q8AUS5 oryza sativ
39	103.5	5.5	691	16	Q9RZS7	Q9RZS7 deinococcus
40	103.5	5.5	1269	17	Q97Z06	Q97Z06 sulfolobus
41	103.5	5.5	2399	16	Q9ZKX9	Q9ZKX9 helicobacte
42	103	5.5	868	17	Q8PZM5	Q8PZM5 methanoscara
43	103	5.5	1754	16	Q8D81	Q8D81 thelamoanear
44	102.5	5.4	1736	5	Q95PH7	Q95PH7 dictyosteli
45	102.5	5.4	1736	5	Q86A19	Q86A19 dictyosteli

ALIGNMENTS

RESULT 1

Q93G67 PRELIMINARY; PRT; 361 AA.
AC Q93G67; 01-DEC-2001 (TREMELrel. 19, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE Minor pilin protein CsaE.
GN CsaE.
CS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=E11881A;
RA Altboum Z.D., Levine M.M., Galen J.E., Barry E.M.;
RT "Isolation and Characterization of ETEC CS4 -imbriae encoding genes,
RL and their expression in Shigella flexneri 2a guaba strain CVD 1204.";
DR EMBL; AF296132; AAK97137.1; -
SQ SEQUENCE 361 AA; 45102 MW; 2E1E74ABD80EB353 CRC64;

Query Match 100.0%; Score 1886; DB 2; Length 361;
Best Local Similarity 100.0%; Pred. No. 6.4e-135;
Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNKILFFTLFFSSVLFTFAVSADKIPGDESITNIFGPRDRNESSPKHNLNNHITAYSE 60

Db 1 MNKILFFTLFFSSVLFTFAVSADKIPGDESITNIFGPRDRNESSPKHNLNNHITAYSE 60

QY 61 SHLYDRMTFLCSSLNLTNGACPTSENSSSVSGETNITLQFTTKRSLSIKELQIKGY 120

Db 61 SHLYDRMTFLCSSLNLTNGACPTSENSSSVSGETNITLQFTTKRSLSIKELQIKGY 120

QY 121 KQLLFKSVNCPSGITLNSAHFNCKNAASGASLYLYIPAGELKNLPGGWDATLKV 180

Db 121 KQLLFKSVNCPSGITLNSAHFNCKNAASGASLYLYIPAGELKNLPGGWDATLKV 180

```
QY 181 RRYSETYGTNTITIKLDKGNIQIWLPOKSDARVDNLNRPRTGGTYIGRNSVDMCFY 240
DB 181 RRYSEYGYTNTITIKLDKGNIQIWLPOKSDARVDNLNRPRTGGTYIGRNSVDMCFY 240
QY 241 DGYSTNSSLEIRFQDNNPKSDGKFLRKINDTKETIAYTSLSLLAGKSLTPNGSLNI 300
DB 241 DGYSTNSSLEIRFQDNNPKSDGKFLRKINDTKETIAYTSLSLLAGKSLTPNGSLNI 300
QY 301 ADAASLETNNRITAVTMPETISVPLVCWPGRLQIDAKVENPEAGQYMGNIINVTFPSSQT 360
DB 301 ADAASLETNNRITAVTMPETISVPLVCWPGRLQIDAKVENPEAGQYMGNIINVTFPSSQT 360
QY 361 L 361
DB 361 L 361

RESULT 2
QY8495 PRELIMINARY; PRT; 363 AA.
AC Q8495;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE CSDD.
GN Escherichia coli.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94344028; PubMed=79:5003;
RA Froehlich B., Karakashian A., Melsen L.R., Wakefield J.C., Scott J.R.;
RT "Cooc and Cood are required for assembly of CSI pill.";
RL Mol. Microbiol. 12:387-401(1994).
DR EMBL; X76908; CAA54230.1; -.
DR PIR; S49539; S49539.
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0004812; P:RNA ligase activity; IEA.
DR GO; GO:0006418; P:amino acid activation; IEA.
DR InterPro; IPR001412; rRNA-synt 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Signal.
FT SIGNAL.
FT CHAIN.
SQ SEQUENCE 363 AA; 40145 MW; 0911574F0406C54A CRC64;

Query Match 50.4%; Score 950.5; DB 2; Length 363;
Best Local Similarity 53.1%; Pred. No. 5.4e-64;
Matches 197; Conservative 54; Mismatches 99; Indels 21; Gaps 11;

QY 3 KILFIPLTFSSVLPFAVSADKIPGDE--SITNIF-GPR-DRNESSPKINLNHITAY 58
DB 2 KKIPIFL----SIIFSVAWSAGRYPTTVGNLTQSFQAPLDRSVQSPINFTNHVAGY 57
QY 59 SESHTLYDRMTFLCLSSHNTINGACPTSENPSSSVS-GETNITLQFTEKRSLIKRELQI 117
DB 58 SLSHSLYDRIVFLCTSSSNPNVNGACPFI---GTSGVQYGTITITLQFTEKRSLIKRNINL 114
QY 118 KGYKQLLFKSVNCPG--LTNSAHFNCNKA-ASGASLYIYPAGELKNLPFGGIWDAT 174
DB 115 AGNKKPIWENQSCDTSNLMVLNKSWSGANGNANGTLLNLYIPAGEINKLPFGGIWEAT 174
QY 175 LKLRVKRYSYSET---YGYTINITIKLDKGN:QIWLPOKSDARVDNLNRPRTGGTYI 230
DB 175 LILRLS-RYGEVSTHYGNYTNTITVDLTDKGN:QVWLPFGHSPNPRVDNLNRPRTGNKYK 233
QY 231 GRNSVDMCFYDGYSTNSSLEIRFQDNNPKSDGKFLRKINDTKETIAYTSLSLLAGKSL 290
DB 234 GSNLDMCFYDGYSTNSDSVVIKQDDNFTYSSEVNLKIG-GTEKLPYAVSLMGEKIF 292
QY 291 TPTNGTSLNIADAASLETNNRITAVTMPETISVPLVCWPGRLQIDAKVENPEAGQYMGNI 350
DB 293 HPVNGQSTINDSSVLETNWNRVTAVAMPENVPVLCWPARLLNADVNADPAGQYSGQI 352
QY 351 NVTFPSSQTL 361
DB 353 YITFTPSVENL 363

RESULT 4
QY8495 PRELIMINARY; PRT; 364 AA.
AC Q8495;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
```

```
QY 181 RRYSETYGTNTITIKLDKGNIQIWLPOKSDARVDNLNRPRTGGTYIGRNSVDMCFY 240
DB 181 RRYSEYGYTNTITIKLDKGNIQIWLPOKSDARVDNLNRPRTGGTYIGRNSVDMCFY 240
QY 241 DGYSTNSSLEIRFQDNNPKSDGKFLRKINDTKETIAYTSLSLLAGKSLTPNGSLNI 300
DB 241 DGYSTNSSLEIRFQDNNPKSDGKFLRKINDTKETIAYTSLSLLAGKSLTPNGSLNI 300
QY 301 ADAASLETNNRITAVTMPETISVPLVCWPGRLQIDAKVENPEAGQYMGNIINVTFPSSQT 360
DB 301 ADAASLETNNRITAVTMPETISVPLVCWPGRLQIDAKVENPEAGQYMGNIINVTFPSSQT 360
QY 361 L 361
DB 361 L 361

RESULT 2
QY8495 PRELIMINARY; PRT; 363 AA.
AC Q8495;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE CSDD.
GN Escherichia coli.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94344028; PubMed=79:5003;
RA Froehlich B., Karakashian A., Melsen L.R., Wakefield J.C., Scott J.R.;
RT "Cooc and Cood are required for assembly of CSI pill.";
RL Mol. Microbiol. 12:387-401(1994).
DR EMBL; X76908; CAA54230.1; -.
DR PIR; S49539; S49539.
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0004812; P:RNA ligase activity; IEA.
DR GO; GO:0006418; P:amino acid activation; IEA.
DR InterPro; IPR001412; rRNA-synt 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Signal.
FT SIGNAL.
FT CHAIN.
SQ SEQUENCE 363 AA; 40145 MW; 0911574F0406C54A CRC64;

Query Match 50.4%; Score 950.5; DB 2; Length 363;
Best Local Similarity 53.1%; Pred. No. 5.4e-64;
Matches 197; Conservative 54; Mismatches 99; Indels 21; Gaps 11;

QY 3 KILFIPLTFSSVLPFAVSADKIPGDE--SITNIF-GPR-DRNESSPKINLNHITAY 58
DB 2 KKIPIFL----SIIFSVAWSAGRYPTTVGNLTQSFQAPLDRSVQSPINFTNHVAGY 57
QY 59 SESHTLYDRMTFLCLSSHNTINGACPTSENPSSSVS-GETNITLQFTEKRSLIKRELQI 117
DB 58 SLSHSLYDRIVFLCTSSSNPNVNGACPFI---GTSGVQYGTITITLQFTEKRSLIKRNINL 114
QY 118 KGYKQLLFKSVNCPG--LTNSAHFNCNKA-ASGASLYIYPAGELKNLPFGGIWDAT 174
DB 115 AGNKKPIWENQSCDTSNLMVLNKSWSGANGNANGTLLNLYIPAGEINKLPFGGIWEAT 174
QY 175 LKLRVKRYSYSET---YGYTINITIKLDKGN:QIWLPOKSDARVDNLNRPRTGGTYI 230
DB 175 LILRLS-RYGEVSTHYGNYTNTITVDLTDKGN:QVWLPFGHSPNPRVDNLNRPRTGNKYK 233
QY 231 GRNSVDMCFYDGYSTNSSLEIRFQDNNPKSDGKFLRKINDTKETIAYTSLSLLAGKSL 290
DB 234 GSNLDMCFYDGYSTNSDSVVIKQDDNFTYSSEVNLKIG-GTEKLPYAVSLMGEKIF 292
QY 291 TPTNGTSLNIADAASLETNNRITAVTMPETISVPLVCWPGRLQIDAKVENPEAGQYMGNI 350
DB 293 HPVNGQSTINDSSVLETNWNRVTAVAMPENVPVLCWPARLLNADVNADPAGQYSGQI 352
QY 351 NVTFPSSQTL 361
DB 353 YITFTPSVENL 363
```

DR	EMBL; AY082893; AAL92876.1; -
DR	InterPro; IPR003016; Lipoyl_BS.
DR	PROSITE; PS00189; LIPOYL_1
DR	SEQUENCE 387 AA; 41653 MW; B4547E146DDAC7B2 CRC64;
QY	Query Match 20.2%; Score 381; DB 2; Length 387;
QY	Best local similarity 32.6%; Pred. No. 8.7e-21;
QY	Matches 116; Conservative 50; Mismatches 146; Indels 44; Gaps 13;
QY	39 RDRNSSPKHNLNNHTAYSEHTL-YDRMTPFLCLSSHNTLNGACPTSENPSSSVSGE 97
Dd	43 RDRGTPGVGDVLIIDFKASGGYDTSYDLKWGRNSWTQSSTDTATGACTV--PVWEKAGT 100
QY	98 TMTLTQFTKRSLIKRELQIK--GYKQ-----LLPKSVNCP--SGLTLNSAHFNON 144
Dd	101 TNKLFSKEENTGATAVLLEGANGQRHRDCTGSVISGVDCPKTDGNRESATIACGD 160
QY	145 KNAASGASILYIPAGELXNLPFGGIWDATLKLRVKRRYSSETGYTYI---NITIKLTDK 201
Dd	161 TVGWGGRVLFVKIPRAELKKLPSGGTWKANRLNKLWSSSTATLAIFKAATLDTTDK 220
QY	202 GNITQLPQFKS-DARVDNLNRTGGTVI-GNSVDMCFDYGYSTNSSSLLEIRPDQN-- 257
Dd	221 NNIQVILPFTSATPTVDLKLNRVANSGRMSGTSNVDMCLYDGSNQSTWFDVASDGLT 280
QY	258 -NPKSDGKYLRKINDT---KEIAYTLUSLLAGKSLCTPTNGSLMIADAASLETWNWR 312
Dd	281 IDRRDGXGTSLT-DKDKSGAYRESDIYAASLTZAGKKIALPNNETVLOG-----VNNSA 335
QY	313 ITAVTMPEISVPVLCPGRLQLDARKVENPE-----AGQFMGNINVTTFPSSOTL 361
Dd	336 GRSVILPG-SAPVICPTPLTL---ETPFQSWWKPGKYSHKLTITFPSSLSL 387
QY	RESULT 6
QY	Q8KQ10 PRELIMINARY; PRT; 387 AA.
ID	Q8KQ10
AC	Q8KQ10;
DT	01-OCT-2002 (TrEMBLrel. 22, Created)
DT	01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	Putative minor pilin and initiator.
DN	CBLD.
OS	Burkholderia cepacia (Pseudomonas cepacia).
OC	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC	Burkholderiaceae; Burkholderia.
OX	NCBI_TaxID=292;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=BC7;
RA	Tomich M., Mohr C.D.;
RT	"Identification and characterization of a locus required for cable
RT	pilus biogenesis in Burkholderia cepacia."
RL	Submitted (May-2002) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AV14293; AAK56040.1; -
DR	InterPro; IPR003016; lipoyl_BS.
DR	PROSITE; PS00189; LIPOYL_1.
QY	SEQUENCE 387 AA; 41805 MW; 39DC45F31BA8CE0C CRC64;
QY	Query Match 19.7%; Score 371; DB 2; Length 387;
QY	Best local similarity 31.7%; Pred. No. 5e-20;
QY	Matches 114; Conservative 50; Mismatches 144; Indels 52; Gaps 13;
QY	39 RDRNSSPKHNLNNHTAYSEHTL-YDRMTPFLCLSSHNTLNGACPTSENPSSSVSGE 97
Dd	43 RDRGTPGVGDVLIIDFKASGGYDTSYDLKWGRNSWTQSSTDTATGACTV--PVWEKAGT 100
QY	98 TMTLTQFTKRSLIKRELQIK--GYKQLPKSYNC-----PSGLTLNSAH 140
Dd	101 TNKLFSKEENTGATAVLLEGANGYQ---RHRCDSGISVGVSARRPNEMAQSATI 156
QY	141 PNCKNDASGASILYIYPAGELKNLPGGIWDATLKLRVKRRYSSETGYTYI---NITIK 197

```

Db 157 AGCDTVGVGDGRVLPVKIPAAELKCLPFGSGTKANLRANLKLWSSSTATTTLA:FKAAITLD 216
QY 198 LTKGNIOIWLPOKXS-DARVDNLNLRPTGGTYI-GRNSVDMCFYDGYSTNSSSLEIRFQ 255
Db 217 TTKXNIOYVLEPFTSATPTVDLKLKPLNGSMGTSNVDMLDYGNSQSTWFDVSAS 276
QY 256 DN---NPKSDGSGFYLRKINDTT---KEIATYLSLLAGKSLTPPTNGTSLNIADAASLET 308
Db 277 DGLTIDRRKQGVSTLDDKSGAVESRDYAASLTYAGKIALPNNETVRLQG-----V 331
QY 309 NNRRITAVTMPEISVPLVCPGRQLQDAKVENP-----AGQYMGNNINVTFTSSQTL 361
Db 332 NNSAGRSVTLPGTSAPVICTPTPLTL-----ETPEFQSVWKRPEKYSKHLTIITFTSSSTSL 387

RESULT 7
Q9XDS1 ID Q9XDS1 PRELIMINARY; PRT; 359 AA.
AC Q9XDS1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE TSAD protein (putative fimbrial protein).
GN TSAD OR TCFD OR STY0348 OR T2547.
OS Salmonella enterica.
OS Salmonella typhi, and
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601, 28901;
RN [1]
SEQUENCE FROM N.A.
RP SPECIES=S.typhi; STRAIN=GIFU10007;
RC Hashimoto Y.;
RA "Identification of a putative fimbrial operon of Salmonella typhi.";
RT Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A.
RP SPECIES=S.typhi; STRAIN=RKS 3333;
RC MEDLINE=99348391; PubMed=10417651;
EX Folkesson A., Advani A., Sukupolvi S., Pfeifer J.D., Normark S.,
RA Lofdahl S.;
RA "Multiple insertions of fimbrial operons correlate with the evolution
RT of Salmonella serovars responsible for human disease.";
RL Mol. Microbiol. 33:612-622(1999).
RN [3]
SEQUENCE FROM N.A.
RP SPECIES=S.typhi; STRAIN=CT18;
RC MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moulle S., O'Gaora P., Parry C.,
RA Quail X., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RA "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [4]
SEQUENCE FROM N.A.
RP SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;
RC MEDLINE=32531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
RL EMBL; AB029403; BAB82272.1; -
DR EMBL; AJ242964; CAB51577.1; -
DR EMBL; AL627266; CAD08773.1; -
DR EMBL; AE016842; AAO70131.1; -
DR PI3; AE0541; AE0541.

```

```

KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 359 AA; 39741 MW; BB88A4E9A62052CE CRC64;
Query Match 14.6%; Score 275.5; DB 16; Length 359;
Best Local Similarity 29.7%; Pred. No. 7.8e-13;
Matches 90; Conservative 48; Mismatches 130; Indels 35; Gaps 11;

QY 70 FICLSHNTLNGACPTSNPSSSVSGSTNITLQTKESLKRBLQIKGYKQLPKSVN 129
Db 73 WVCRRNRNENEGACETHLVWVAFGAYSKRLRFREQTSHAETL-----ILLGSVR 131
QY 133 --CPSSLTNGSAHFHONKNAAS--GASLYLIPAGELKNLPFGGIWDATLKL-RVKRRY 183
Db 132 DACYTSV-----INMNAACQGRSLKRLPSEELAKIPTSGTWKATLVLDLWGG 183
QY 184 SETYGTYTITIKLTD--KGNIQIWLPOF-KSDARVDNLNLRPTGGTYIGRNSVDMCFY 240
Db 184 DPLGTSTTIDILNVDHFAENAAIYFPQGTATPRVDNLNLRHMASQMSGRANLDMCLY 243
QY 241 DGYSTNSSLEIRFOONPKSDGKFLYLRKINDTKEIATYLSLLAGKSLTPPTNGTSLNI 300
Db 244 DG-GVKARSLOWKIEGSKSGTGFQVKSADSADT--IDYAVSMYNGRSIPVTRGVFSL 300
QY 301 ADAASLETNNRITAVTMPEISVPLVCPGRQLQDAK---VENPAGQWGMNINVTFTPS 357
Db 301 DNVDKAAIR-----PVLPQGRQAVRCVPLTLTTPPNIREKSGEYQGITLVMLMG 355
QY 358 SQT 360
Db 356 TQT 358

RESULT 8
Q83JU1 ID Q83JU1 PRELIMINARY; PRT; 245 AA.
AC Q83JU1;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Orf, partial conserved hypothetical protein.
GN SP3134.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=301 / Serotype 2a;
RC MEDLINE=22272436; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng E., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RA "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441(2002).
DR EMBL; AE015325; AAN44606.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 245 AA; 27602 MW; 4D27007AE2F0B1A3 CRC64;

Query Match 14.4%; Score 272.5; DB 16; Length 245;
Best Local Similarity 32.0%; Pred. No. 8.1e-13;
Matches 73; Conservative 37; Mismatches 89; Indels 29; Gaps 9;

QY 153 LYLIPAGELKNLPFGGIWDATLKLVRKRYSTGYTYITIKLTKGNIOIWLPOF- 211
Db 27 LTYLQTEKKNLPICGVW---KGVKLNHNSPAQDYFANITLTDPHVIDVFPEPA 82
QY 212 KSDARVDNLNLRPTG---GGTYI-GRNSVDMCFYDGYSTNSSSLEIRFQDN---PKSDG 263
Db 83 HATPRQLDLHPTGVSNGSNYAQDLTMDLWCLYDGFNGNAISYEIMLKDRGPAAGRRDG 142

```

```

QY 264 KYLRK---INDDTKETAYTUSLLACKSLTPTNGTSLNIADAASL---ETNWNRIYAV 316
DB 143 YFSIYRQGGTTDEGERIDYR-----KYNFETGGQIDVRNENVMWNSINLKRVRPV 196
QY 317 TMEPISVPVLCWPGRLQIDA---KVENPEAGYMGNIINVTPESSQTL 361
DB 197 VLPGRYAVMCPVPTLTLAVDRFSVNDKQAGYMGKLSVIFTPSLPTI 244

RESULT 9
QYUBH8 PRELIMINARY; PRT; 210 AA.
AC QYUBH8;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN S3341
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T."
RL Infect. Immun. 71:2775-2786(2003).
DR EMBL; A2016989; AAP18419.1; --
KW Hypothetical protein.
SQ SEQUENCE 210 AA; 23516 MW; BBD109DC114A2486 CRC64;

Query Match 13.6%; Score 256.5; DB 16; Length 210;
Best Local Similarity 31.5%; Pred. No. 1.1e-11;
Matches 69; Conservative 36; Mismatches 85; Indels 29; Gaps 9;

QY 162 LKNLPGGIWDATLKLVRKRRSETGTYYTINIKLTKGNQIWLQPF-KSDASVDLN 220
DB 1 MNKLPGGVW---KGVKLHNSPPAQDYFANITLTPNH-DVFPFBFAHATPVQLD 56
QY 221 LRPTG---GQTYI-GRNSVDMCFYDGYNSSSLSIRFQDN---PKSDGKFYLRK--- 269
DB 57 LRPTGSVNSVAQDLTMDCLYDFGNAGISYRMLKDECRPAAGRDGYFSYRQGG 116
QY 270 -INDDTKETAYTUSLLACKSLTPTNGTSLNIADAASL---ETNWNRIYAVTMEPISVPV 325
DB 117 TTTDEGERIDYR-----KYNFETGGQIDVRNENVMWNSINLKRVRPVLPGRYAV 170
QY 326 LCPGRLQIDA---KVENPEAGYMGNIINVTPESSQTL 361
DB 171 MCVPTLTLAVDRFSVNDKQAGYMGKLSVIFTPSLPTI 209

RESULT 10
QYUBH8 PRELIMINARY; PRT; 1801 AA.
AC QYUBH8;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hemolysin.
GN FN0291.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Ponstein M., Kyrides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586."
RL J. Bacteriol. 184:2005-2018(2002).
DR EMBL; A2010541; AAL94497.1; --
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:protease activity; IEA.
DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008638; Haemaggl.
DR InterPro; IPR000508; Peptidase_S26.
DR InterPro; IPR001452; SH3.
DR Pfam; P705860; Haemaggl_act; 1.
DR Pfam; P700018; SH3; 1.
DR PROSITE; PS00761; SPASE_I_3; 1.
KW Complete proteome.
SQ SEQUENCE 1881 AA; 204375 MW; D635156A4EFA4877 CRC64;

Query Watch 6.7%; Score 127; DB 16; Length 1881;
Best Local Similarity 23.1%; Pred. No. 1.2; Indels 122; Gaps 19;
Matches 96; Conservative 45; Mismatches 122; Gaps 19;

QY 1 MNKLIPIFTLFSVLTFAV-----SADKIPGDESITNIFGPRDRNESSPKHN- 49
DB 5 LKLLIAIFMLHIISLADGIVPDNSASKNLQVKAANGVPLVNEAP-DNNGTS--HKV 61
QY 50 -----LNNHITAYSESHLYDRMTFLCLSSHNTILGACETSENPSSSYSG-- 96
DB 62 YKDYNVDRGAILNN-----SKLTNSQLGGII 89
QY 97 ETNITLQTEKPSLIKREL-----QIKGYKOLLFKSVNC-----PSGLTLGAHF--NCN 144
DB 90 YGNPLQNSKEASTINEVGVNKRIGIQIVGRANYILANPNIANGACFINTGN 149
QY 145 KNAAGSASLYLIP-----AGEIKNLPFGGIWDATLKLVRKRRSETGTYYTINIK 197
DB 150 VFTTGRGNLNPKEKMEIAG--KGLDLRNINKAELARVAELSAPIYGGVEVNLKG 207
QY 198 LNDKNIQIWLQPFKSDARVDNLNLRPTGGTIGNSVDMCFYDGYNSSSLSIRFQ-- 255
DB 208 NQKSN---KPEYALDAR-----ALGSIYAGRINI-----IVNEDGVGVKTQAP 248
QY 256 -----DNPKSDGKFYLR-----KINDDTKEIAYTUSLLACKSLTPTNGTSLNIA 301
DB 249 MYATKGVVISKGVLYKDTQAKRDIKISSTETEIG---SKLLAENAINIKSGK--SN-- 303
QY 302 DAASLETNWNRI 314
DB 304 -SQQIRAN--NNIT 314

RESULT 11
QYUBH8 PRELIMINARY; PRT; 700 AA.
AC QYUBH8;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Putative polymorphic membrane protein (Fragment).
OS Chlamydia psittaci (Chlamydia psittaci).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=POS;
RA Laroucau K., Souriau A., Rodolakis A.;
RT "Isolation of a new pmp sequence and evidence of pmp polymorphism in
RT serotype-1 Chlamydia psittaci strains."

```

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF243415; AAL36959.1; --
 DR InterPro; IPR003368; Chlamydia_PMP.
 DR Pfam; PF02415; Chlamydia_PMP; 1.
 DR TIGRFAMs; TIGR01376; POMP_repeat; 3.
 FT NON_TER 700 700
 SQ SEQUENCE 700 AA; 74200 MW; 2C38E504AF48D1D1 CRC64;

Query Match 6.1%; Score 115.5; DB 2; Length 700;
 Best Local Similarity 18.3%; Pred. No. 2.5; Indels 139; Gaps 18;
 Matches 73; Conservative 63; Mismatches 124;

QY 42 NESSPKHILNNHITA-----YSSHTLYDRMTFLCL---SHNTLNGAC----- 53
 DB 27 NEELTSSDSYNGNVISDEFEVXETTSAGIATCEGNVCISYAGKOSPLNKSCFSETTENLS 86
 QY 84 -----PTSENPSSSSVSGETNITLQFTKESLKLRELOIKGYKQLLPKSVN 129
 DB 87 FIGNGYTLCFDNITTTASNGAINVSGD-----QKTLNVSGFS--LFSCAH 130
 QY 130 CPSGLT-----LNSAHF--NCNK-----NAASGASLYLYIP-----A 159
 DB 131 CPPTGTGYGAIOATKGVSTPSGNKLIFDNCCSGEGGAIKCATGSAEKLKENSIVVFS 190
 QY 160 GELKMLPFGGIMDAILKLVKRYSEYTYGTITINIKLTKDGNIQIWLPOPKSDARVDL 219
 DB 191 GNSQKKGAIY--TKLITADGPTLPSNNSVAS--SPKGG-AICLDGTSSECSLTA 244
 QY 220 NL-----RPTGGTYVIGRNVDM-----CFYDGYSTNSSLR 251
 DB 245 NLGDITFDGNKVIKINGGSGSTVRNAIDLGGGKFTKLNKAKGEGFIFHPDIANTGGSTE 304
 QY 252 IRFDNNPKSD---CK--FYLRKINDDTKEIAYTL-----SILLA 286
 DB 305 IEL--NATESDITTYTKIVFSGSKLDEEKTVPDNLKSVFKQPLKICAGSLVLKDGVTLE 362
 QY 287 GKSLTPTNGTSLNIADAASLETNNRITAVTMEISVPV 325
 DB 363 AKKITOTKGVTVVGLGTTTQTSSSGETITLNLINI 401

RESULT 12

P711132 PRELIMINARY; PRT; 847 AA.
 AC P71132; 1997 (TrEMBLrel. 02, Created)
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE POMP91A.
 OS Chlamydia abortus.
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydiaophila.
 OX NCBI_TaxID=83555;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=S26/3;
 RX MEDLINE=96406378; PubMed=6810511;
 RA Longbottom D., Russell M., Jones G.E., Lainson F.A., Herring A.J.;
 RT "Identification of a multigene family coding for the 90 kDa proteins
 of the ovine abortion subtype of Chlamydia psittaci";
 RL FEMS Microbiol. Lett. 142:277-281 (1996).
 [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=S26/3;
 RX MEDLINE=98187897; PubMed=9529048;
 RA Longbottom D., Russell M., Dunbar S.M., Jones G.E., Herring A.J.;
 RT "Molecular cloning and characterization of the genes coding for the
 highly immunogenic cluster of 90-kilodalton envelope proteins from the
 Chlamydia psittaci subtype that causes abortion in sheep";
 RL Infect. Immun. 66:1317-1324 (1998).
 RL EMBL; U65942; AAC15921.1; --
 DR InterPro; IPR006315; Autotransport.
 DR InterPro; IPR003368; Chlamydia_PMP.
 DR Pfam; PF02415; Chlamydia_PMP; 1.

DR TIGRFAMs; TIGR01414; autotrans_bar1; 1.
 DR TIGRFAMs; TIGR01376; POMP_repeat; 3.
 SQ SEQUENCE 847 AA; 90694 MW; 754C958E7F11179E CRC64;

Query Match 6.1%; Score 115.5; DB 2; Length 847;
 Best Local Similarity 21.6%; Pred. No. 3.2; Indels 131; Gaps 21;
 Matches 94; Conservative 58; Mismatches 153;

QY 10 LFFSVLFTFAVSADKIFGDSITNIFQPRDRNESPSPKHNILNNHITAYSSHTLYDRMT 69
 DB 9 LSSSLIVSNLSYSE--PDQKTLTSAHSGVNGTNSP-----FNPLSTSNNGTIYCTG 62
 QY 70 FLCLs-----SHNTLNGACP-----SENPSSSSVSGETNITL 102
 DB 63 NICIAYAGLDGSLSSSCFTDTAGNLFLGNGYTLCFDNITTOSSHHPGAISSVG-TNKT- 120
 QY 103 QFTKRSLLIKRELQIKGYKQLLFKSWPCSPGLTNSA-----HFNCKN 145
 DB 121 -----LDISGFS--LFSCAYCPGATGYGAIKAVGNTTIKONSSLVPHKNCST 166
 QY 145 NAASGASLYLYIPAGELK-----NLPP-----GGTWDATLKLVRKRYSETVGTIT 192
 DB 167 GEGGAIQCKKASSSEAELEKTIENNQLVFAENSSSSSGGAIYAD-KLTISSGGTFLFSNNSV 225
 QY 193 NIT-----IKLTDKGNIQIWLPOPKSDARVDL--LRPTGGTYVIGRNVDM----- 237
 DB 225 SASPKGGAICIKDSGG-ECSLTAGLGGITFDGNKLIKINGGSPVTRNSIDLSSGSKFT 284
 QY 238 -----CFYDGYV-TNSSLEIRFQDNNPKSDGK--FYLRKINDDTKEIAYTL----- 281
 DB 285 KLNKAKGFIPIFYDPIPTGGGSDLNINKQDITVDYT-GKIVFSGERLSDSEKVAANLKSD 343
 QY 282 -----SILLAG-----KSLTPTNGTSLNIADAASLET--NWNRITAVTMEISV 323
 DB 344 FKQPLKIGSGSLIKDGVLTETKSTQTGTGATVMDLGTTLQTPSSGGETITLNLINIV 403
 QY 324 PVLCPGRQLDAKVE 339
 DB 404 ASLGGGVAPDPAKVE 419

RESULT 13

Q8Q7H7 PRELIMINARY; PRT; 865 AA.
 AC Q8Q7H7;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Envelope glycoprotein (Fragment).
 GN ENV.
 OS Human immunodeficiency virus 1.
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11676;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=99CMA304;
 RX MEDLINE=21849375; PubMed=1183674;
 RA Yamaguchi J., Vallari A.S., Swanson P., Rodelle P., Kaptue L.,
 RA Nansop C., Zekeng L., Gurtler L.G., Devare S.G., Brennan C.A.;
 RT "Evaluation of HIV Type 1 Group O Isolates: Identification of Five
 Phylogenetic Clusters";
 RL AIDS Res. Hum. Retroviruses 18:269-282 (2002).
 DR EMBL; AF383244; AAL98866.1; --
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR000328; Env-Gr41.
 DR InterPro; IPR000777; GP120.
 DR Pfam; PF00516; GP120; 1.
 DR Pfam; PF00517; GP41; 1.
 DR ALDS; Coat protein; Glycoprotein; Polyprotein; Transmembrane.
 KW ALDS; Coat protein; Glycoprotein; Polyprotein; Transmembrane.
 FT NON_TER 1 1

SQ SEQUENCE 865 AA; 97118 MW; 996883862AA31ACA CRC64;
Query Match 6.0%; Score 114; DB 15; Length 865;
Best Local Similarity 21.5%; Pred. No. 4.2;
Matches 93; Conservative 56; Mismatches 150; Indels 134; Gaps 22;
QY 21 VSADKTPGESITNFGPRDRNESSPKNINLNH-----ITAYSSHLYDRMTFL 71
DB 75 VPTDTPPEYPLHN-----TDKFIWENYVYKQMGDIIDWEOSLKPCKVMTFL 125
QY 72 C-SSNMT-NGACPTSENSSSVS-----GETNITLQF-EKSLIKRELQIKYKOLLEK 126
DB 126 CVMQNTSLN---DKSNFTSSPENLMKKCFNVTWVKQKKAALFYVSDMLKLDK 182
QY 127 SYNCPSGLTNGAHFNCKNAASGA-----SYLYIPAG-----ELKNLPFGG---- 169
DB 183 NITNTMTYLT---NONSFTISOACPKVSFEPIPIHYCAPAGAYAFKCNNEFNSTGTC 238
QY 170 -----INDATL---KLV-KRYSETYG-TYTNITIKLTDK--G 202
DB 239 KNITVTCTHGKPTVSTQILNGTILSKGIRIMTKNISDNVNIIVTLASTLKITCRPG 298
QY 203 NIQIWLPOKSDARVDNLNRPCTGGTYIGNSVDMCFYDGYSTNSGSLRIFODNNPKSD 352
DB 299 NMTVQOVSTGPNWYSDLR-----GTERNTSRVAYCEY-----NSIDWE-----RTLKQT 344
QY 263 GKPYLRKINDYKXIATYLSLLAG-----KSL-----TPTNGTSL 298
DB 345 AERFLSLVN-NTKKVDMTFNSGGDPVANLHFNCHGEFFYCNSTSLFNFTSCNGSTC 403
QY 299 NIADAASLETNWRI-----TATMPEISVPVLQWPGRLDQAKVNPBA 343
DB 404 NVTKSN-NSTNTRIPCLRQVVKVMTQGGSGLYAPPINRLTGMNITGLMTQMDQP-- 460
QY 344 GQYMGHNITFTP 356
DB 461 ---WNHNSATFRP 470
RESULT 14
Q81521 PRELIMINARY; PRT; 2646 AA.
AC Q81521
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Erythrocyte membrane protein 1 (pEMP1).
GN PF01950W.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7; PubMed=12358864;
RX MEDLINE=22255705; PubMed=12358864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W., K.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Rene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Perteu M., Allen J., Selegut J., Haft D., Mather W.M., Vaidya A.B.,
RA Martin D.N.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
ET "Genome sequence of the human malaria parasite Plasmodium falciparum.";
RL Nature 415:498-511(2002).
DR EMBL; AE014850; AAN36476.1; -.
DR GO; GO:0005539; F:glycosaminoglycan binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR InterPro; IPR004256; PfEMP.
DR Pfam; PF03011; PfEMP; 2.
SQ SEQUENCE 2646 AA; 300284 MW; C91D25F3DE78717 CRC64;

Query Match 6.0%; Score 113; DB 5; Length 2646;
Best Local Similarity 24.8%; Pred. No. 21;
Matches 77; Conservative 37; Mismatches 128; Indels 68; Gaps 15;
QY 23 GDESITNFGPRDRNESSPKNINLNHITAYSSHLYDRMTFLCLSSNTLNGACPTG- 86
DB 695 GDS--TAIQIGIKTKLEKKKKGAN-----ISEEKTILDEFNLHEIKDAETCKNCEPRXF 747
QY 87 ENPSSSVSGEINILQFTEKESLIKRELQIKYKOLLFPKSYNCPGSLTNSAHFNCKN 146
DB 743 KNPCSGDTSN--KQVEAVANTVAQLQGAQKOL-----HNGSRN 789
QY 147 AASGASLYIIPAGELKNLPFGGINDATLKLVRKRYSETYTYTINIKLTDKNI-- 204
DB 790 ALKGNIQAKINNGKPN-----PLTDACQITKXHSNGK-----DSNNPCNKNRLK 838
QY 205 --QIWLPOKSD-ARVDL-----NLRPTGGTYIGNSVDMCFYD--YSTNS 247
DB 839 IQQVW--SIKNDTSYTDVTPPRQHVCTSNLEKLYASVIGSNVNDKFLVEVLHAAS 896
QY 248 SLEIRFQNNPKSDGKFIYRKINDT-KEIAYT---LSLLAGSLTPTNGTSLNADA 303
DB 897 EAEFIKKYNEKQNDCKGLRKQATTCRAIRYSFADIGDIINGKOLWDDNN-----DA 950
QY 304 ASLETNWNEI 313
DB 951 KSLQTNLKA 960
RESULT 15
Q8XNW2 PRELIMINARY; PRT; 743 AA.
AC Q8XNW2
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein CP0220.
GN CP0220C.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX MEDLINE=21664373; PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
ET "Complete genome sequence of Clostridium perfringens, an anaerobic flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL; AP003185; BAB79526.1; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR006635; NEA_transpt.
DR SMART; SM00725; NEAT; 4.
DR TIGRfam; TIGR01167; LPXTC_anchor; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 743 AA; 81149 MW; A7C451E39C5A5E545 CRC64;
Query Match 5.9%; Score 111.5; DB 16; Length 743;
Best Local Similarity 18.8%; Pred. No. 5.4;
Matches 64; Conservative 66; Mismatches 115; Indels 95; Gaps 13;
QY 7 IFILFFSSVLTFAVADKTPGESITNFGPRDRNES-----SPKNI-LNNHITAYSES 61
DB 88 LMTVVFNSSLYGFNMNIEVSGGAEKAIENKDKKSIFFVPSPTCKVIGLFTMMGRK 147
QY 62 HTLYDRMTFLCLSSHNTLN--GACPTSENSSSV-----SG-----ETNITLQ 103
DB 148 VELF-----LVNDMTVNLLDAPTINNADISVTQGDADILLGVIQTKSDSNLKE 201
QY 104 PTEKESLIK-RELQIKYKOLLFPKSYNCPGSLTNSAHFNCKNKAASGASLYIYPAGEL 162

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

om protein - protein search, using sw model

Run on: May 6, 2004, 10:11:46 ; Search time 18 Seconds
(without alignments)
1044.296 Million cell updates/sec

Title: US-09-839-894-10
Perfect score: 1886
Sequence: 1 MNKILFIFLTFSSVLTFA.....EAGQTMGNINVTFTSSQTL 361

Scoring table: ELOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1558.5	82.6	360	CFAB_ECOLI	P25734 escherichia
2	120	6.4	2650	YEEU_E057	Q8x8v7 escherichia
3	114	6.0	650	SAG1_YEAST	P20840 saccharomyc
4	109	5.8	2358	YEEJ_ECOLI	P76347 escherichia
5	108	5.7	524	GAS3_YEAST	Q03655 saccharomyc
6	107	5.7	304	YDEQ_ECOLI	P77588 escherichia
7	100	5.3	737	ALYS_ENTFA	P37710 enterococcu
8	99	5.2	802	SYFB_LASIN	Q92016 listeria in
9	99	5.2	1849	IGA4_HABIN	P45386 haemophilus
10	98.5	5.2	608	YD93_METJA	Q58788 methanococc
11	98	5.2	758	SP21_YEAST	P35209 saccharomyc
12	98	5.2	1157	N133_YEAST	P36161 saccharomyc
13	95.5	5.1	713	CDG2_PABMA	P31833 paenibacill
14	95	5.0	413	RPLM_YEAST	P30775 saccharomyc
15	95	5.0	2334	WAPA_BACSU	Q07833 bacillus su
16	94.5	5.0	959	N100_YEAST	Q02629 saccharomyc
17	94	5.0	471	LEU2_BUCRP	P48573 buchnera ap
18	93.5	5.0	412	THBG_SHEEP	P50450 ovis aries
19	93	4.9	1167	CLAA_BACTU	P56956 bacillus th
20	93	4.9	1773	DIP2_DROME	Q9W099 drosophila c
21	92.5	4.9	320	CYF_CYACA	Q9C184 cyanidium c
22	92.5	4.9	544	AGM1_CANAL	Q9P4V2 candida alb
23	92.5	4.9	1061	OAR_MYXA	P38370 myxococcus
24	92.5	4.9	2278	PAB1_YEAST	P34756 saccharomyc
25	91.5	4.9	500	SVK_BUCRP	Q89ac5 buchnera ap
26	91.5	4.9	869	CFAC_ECOLI	P25733 escherichia
27	91.5	4.9	1783	Y468_MYCE	Q49460 mycoplasma
28	91	4.8	802	SYFB_LISNO	Q9y7q1 listeria mo
29	90.5	4.8	309	ELTB_CLOPE	P01558 clostridium
30	90.5	4.8	863	MCMA_XENLA	P30664 xenopus lae
31	90.5	4.8	1019	ENTK_HUMAN	P88073 homo sapien
32	90.5	4.8	1328	HUS2_SCHPO	Q09811 schizosacch
33	90.5	4.8	2193	POLG_HE71M	Q66479 h. genome po

ALIGNMENTS

RESULT 1

```
CFAB_ECOLI
ID CFAB_ECOLI STANDARD; PRT; 360 AA.
AC P25734;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE CFA/I fimbrial subunit B (Colonization factor antigen I subunit B).
GN CFAE.
OS Escherichia coli.
OG Plasmid NTP513.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID:562;
[1]
SEQUENCE FROM N.A.
RC STRAIN=Enterotoxigenic;
RX MEDLINE=89330163; PubMed=2569152;
RA Hamers A.M., Pel H.J., Willsnaw G.A., Kusters J.G.,
RA van der Zeijst B.A.M., Gastra W.,
RT "The nucleotide sequence of the first two genes of the CFA/I fimbrial
RL operon of human enterotoxigenic Escherichia coli.";
RL Microb. Pathog. 6:297-309(1989).
[2]
SEQUENCE FROM N.A.
RX MEDLINE=92329981; PubMed=1352712;
RA Jordi E.J.A.M., Willsnaw G.A., van der Zeijst B.A.M., Gastra W.;
RT "The complete nucleotide sequence of region I of the CFA/I fimbrial
RL operon of human enterotoxigenic Escherichia coli.";
RL DNA Seq. 2:257-263(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMEL; M55661; AAC41417.1; --
DR PIR; D56617; D56617.
DR InterPro; IPR008957; FN III-like.
KW Antigen; Fimbria; Plasmid.
SQ SEQUENCE 360 AA; 39903 MW; 691509B63A8E69CE CRC64;
```

Query Match 82.6%; Score 1558.5; DB 1; Length 360;
Best Loca. Similarity 81.2%; Pred. No. 3.6e-111;
Matches 293; Conservative 31; Mismatches 36; Indels 1; Gaps 1;
Qy 1 MNKILFIFLTFSSVLTFAVSKDKIPDPSINIFGPRDRNSSPKHNLNNHITAYSE 60
Db 1 MNKILFIFLTFSSVLTFAVSKDKIPDPSINIFGPRDRNSSPKHNLNNHITAYSE 60
Qy 61 SHTYSDMTFLCLSSSHNTNGACPTSPNSSSVSGETNITLQFTKRSIKRELQIKGY 120
Db 61 SHLYDMSFLCLSSQNTNGACPSAPGATIDGTNITLQFTKRSIKRELQIKGY 120

```

QY 121 KQLLFKSNVCPGLTNSAHFNCKNAAGSASLYLIPAGELKNLPFGGIDATLKLRYK 180
DB 121 KQFLFKNANCPKALNSSHFCNREQACAGATLSLIPAGELKNLPFGGWNALKLRYK 180
QY 181 RRYSEYGVYITINIKLTDKGNQIWLQPKQSDARVDLNLRTGGTYIGRNSVDMCFY 240
DB 181 RRYDITYGVYITINIVLTDKGNQIWLQPKQSNARVDLNLRTGGTYIGRNSVDMCFY 240
QY 241 DGYSTNSSSLRIPQNNPKSDGKFLYLRKINDTKELIATLSLLIACKSLTPTNGTSLMI 300
DB 241 DGYSTNSSSLRIPQNNPKSDGKFLYLRKINDTKELIATLSLLIACKSLTPTNGTSLMI 300
QY 301 ADAASLETNWRITAYTMPPEISVPLVCPGRLQDQKVENPEAGQYMGINIVTFTPSOT 360
DB 302 -NTASLETNWRITAYTMPPEISVPLVCPGRLQDQKVENPEAGQYMGINIVTFTPSOT 359
QY 361 L 361
DB 360 L 360

RESULT 2
YEEU ECO57 STANDARD; PRT; 2660 AA.
AC Q8X6V7; O8X2B9; O8X2C0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Hypothetical protein yeeU.
GN 23135 OR EGS2775/EGS2776.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83324;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; EpubMed=11256551;
RA Borna N.T., Plunkett G., Zil, Buriand V., Mau B., Glaesner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RI "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RMD 050952;
RX MEDLINE=2115231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
RA Kubara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -!- SIMILARITY: Contains 16 Big-1 domains.
CC -!- SIMILARITY: Belongs to the intimin/invasin family.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC frameshift in position 1315.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by ard for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE005423; AAG57041.1; -
DR EMBL; AP002559; BAB36190.1; ALT_FRAME.

```

```

DR EMBL; AP002559; BAB36199.1; ALT_FRAME.
DR InterPro; IPR003344; Big_1.
DR InterPro; IPR003355; Intimin.
DR InterPro; IPR008964; Invasin_intimin.
DR InterPro; IPR000601; PKD.
DR Pfam; PF02369; Big_1; 16.
DR PRINTS; PR01369; INTIMIN.
DR SMART; SM00634; BID_1; 16.
DR SMART; SM00089; PKD; 8.
KW Hypothetical protein; Repeat; Complete proteome.
FT DOMAIN 738 834
FT DOMAIN 840 929
FT DOMAIN 931 1033
FT DOMAIN 1042 1132
FT DOMAIN 1134 1236
FT DOMAIN 1245 1335
FT DOMAIN 1337 1439
FT DOMAIN 1448 1539
FT DOMAIN 1548 1652
FT DOMAIN 1653 1750
FT DOMAIN 1751 1855
FT DOMAIN 1856 1957
FT DOMAIN 1963 2056
FT DOMAIN 2065 2156
FT DOMAIN 2157 2252
FT DOMAIN 2254 2355
SQ SEQUENCE 2660 AA; 280062 MW; 01EB92A08F5C09D2 CRC64;

Query Match 6.4%; Score 120; DB 1; Length 2660;
Best Local Similarity 20.4%; Pred. No. 0.79; Mismatches 156; Indels 90; Gaps 15;
Matches 80; Conservative 66;

QY 13 SSVLFPTFAVSADKIPGDESITNIFGPRDRNESSPKENILNNHITAY---SESITLDRMT 69
DB 737 SAKIALLSASNNGLANENAA--VSVVADEGS---NPINDHTVTPAVLSSGSA--FNNQN 793
QY 70 FLCLSHNTINGACPTSENPSSSVSGEITNIIQTFEKSLLI-----K 112
DB 794 ----TAKTDVNGLIA--TFDLKSKQEDNTVEVLENGVKQTLIVSFVCDSTQAQVDLQSK 848
QY 113 REIQIKGY-KQLLFKSNVCPGLTNSAHFNCKNAAGSASLYLIPAGELKNLPFGGIW 171
DB 849 NEVVDAGNSATWATVVRDAKGNLNDVKVTNVNSAAKLSQTEVNSHD-----GIA 901
QY 172 DATLKLVRKRYSEYGVYITINIKLTDKGNQIWLQPKQSDARVDLNLRTGGTYIG 231
DB 902 TALT-----SLKNGDYTVTASVSSGSAQNVIFIGQSTAALPLSV-PSGDITV-- 951
QY 232 RNSVDMCFYGVYSTNSSSLR-----IRFQDNNPKSDGKFLYLRKINDTKELIATLSLLIAG 287
DB 952 -----TNTAPLHMTATLQDKNGNPLKDKBITFVSFVND-----VASRFSISNSG 994
QY 288 KSLTPTNGTSL-----NIADAASLETNWRITAYTMPPEISVPLVLC 327
DB 995 KGMTDSNGTATIASLTGLTGHMTATLRLANSVSDTQPMTFVADKQDAVVVLCTSKAEII 1054
QY 328 WPG--RLQLDQKVENPEAGQYMGINIVTFTPS 357
DB 1055 GNGVDEITLTATVKDP-FDNVKNLSVVFRTS 1085

RESULT 3
SAGI YEAST STANDARD; PRT; 650 AA.
ID SAGI YEAST
AC P20840.
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Alpha-agglutinin precursor (A3-alpha-1).
GN SAG1 OR AGAL1 OR YJR004C OR J1418.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

```


RESULT 4

YEEJ ECOLI
 ID YEEJ ECOLI STANDARD; PRT; 2358 AA.
 AC P76347; P94750;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein yeeJ.
 GN YEEJ OR B1978.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97251358; PubMed=9278503;
 RA Blattner F.A.; Plunkett G. III, Bloch C.A., Perna N.T., Burlard V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 279:1453-1474 (1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97251358; PubMed=9087040;
 RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,
 RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
 RA Saito N., Sempel G., Seki Y., Sivasubram S., Tagami H.,
 RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horuchi T.;
 RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 40.1-50.0 min region on the linkage map.";
 RL DNA Res. 3:379-392 (1996).
 CC -!- SIMILARITY: Contains 13 Big-1 domains.
 CC -!- SIMILARITY: Belongs to the intimin/invasin family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE000289; AAC75042.1; ALT_INIT.
 DR EMBL; D90837; BAA15800.1; ...
 DR EMBL; D90836; BAA15799.1; ALT_INIT.
 DR EcGene; EG13378; yeeJ.
 DR InterPro; IPR003344; Big_1.
 DR InterPro; IPR003535; Intimin.
 DR InterPro; IPR008964; Invasin_intimin.
 DR InterPro; IPR002482; lysM.
 DR InterPro; IPR000601; PKD.
 DR Pfam; PF02369; Big_1; 13.
 DR PRINTS; PR01369; INTIMIN.
 DR SMART; SM00634; BID_1; 13.
 DR SMART; SM00257; lysM; 1.
 DR SMART; SM00089; PKD; 6.
 KW Hypothetical protein; Repeat; Complete proteome.
 FT DOMAIN 738 834
 FT DOMAIN 834 931
 FT DOMAIN 931 931
 FT DOMAIN 932 1033
 FT DOMAIN 1042 1137
 FT DOMAIN 1146 1237
 FT DOMAIN 1246 1350
 FT DOMAIN 1351 1448
 FT DOMAIN 1449 1553
 FT DOMAIN 1554 1655

FT DOMAIN 1661 1754
 FT DOMAIN 1763 1853
 FT DOMAIN 1855 1950
 FT DOMAIN 1952 2053
 FT COMPACT 105 105
 SQ SEQUENCE 2358 AA; 248599 MW; 2322497503P631ED CRC64;
 Query Match 5.8%; Score 109; DB 1; Length 2358;
 Best Local Similarity 20.0%; Pred. No. 4.6;
 Matches 78; Conservative 65; Mismatches 157; Indels 90; Gaps 15;
 QY 13 SSVLFTFAVSADKIPGDSITNIFGRDRNESSPGHNLNKHITAY---SESHLYDRMT 69
 DB 737 SAKIATLSASNNGLANEAANTVSVNVADEGS---NPINDHTVTFAVLSGSATSFNNQN 793
 QY 70 FLCLSSHNILNACPTSENSSSVSGETNITLQPTKRSLLI-----K 112
 DB 794 ----TAKTDVNGLA-TFDLKSSKQEDNTVEVLENGVKTLLIVSFWGDSSTAQVDLQSKS 848
 QY 113 RELQIKGYKQL-LFKSVNCPGSLTNSAHFNCKNAASGASLYLIPAGELKNLPFGGIW 171
 DB 849 NEVVADGNDSTMTATVDAKGNLLNDVMVTFNVASAKLSQTEVNGHD-----GIA 901
 QY 172 DATLKRVKRRYSEYCYTYTINITIKLDKGNIQIWLQPKSDARVDNLNLRPTGGTYIG 231
 DB 903 TAITLT-----SLXNGDYRVTA SVSSGSQANQQVNFIDQQTAAITLSV-PSGDIIV-- 351
 QY 232 RNSVDMCFYDGYSTNSS----SLEIRFQDNNEPKSDGKFLRKINDDTKEIAYTLLSLLAG 287
 DB 952 -----TNTAQYMTATQDKNGNPLKKEITFSVPND-----VASKFSISNGG 994
 QY 289 KSLTPTNGTSL-----NIADAASLETWNRAITATMTMEISVPLVC 327
 DB 995 KGMTSDSNGVAIASLTGTLAGTEIMARLANSVSDAQPMTFVADKDRVVVLTQSKAEII 1054
 QY 328 WPG--RLQILDAXVNPACQYMGNIIVTPT 355
 DB 1055 GNGVDETTLTATVKDP-SNHPVAGITVNT 1083
 RESULT 5
 GAS3_YEAST STANDARD; PRT; 524 AA.
 ID GAS3_YEAST
 AC Q03655;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Glycolipid anchored surface protein 3 precursor.
 GN GAS3 OR YMR215W OR YMR261.09.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes;
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RX MEDLINE=97313268; PubMed=9169872;
 RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
 RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
 RA Jørgensen K., Lyle G., Moule S., Ogilvie C., Pearson D., Rajandream M.A.,
 RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome
 XIII.";
 RL Nature 387:90-93 (1997).
 RN [2]
 RP IDENTIFICATION.
 RX MEDLINE=20529944; PubMed=11079560;
 RA Pardo M., Ward M., Bairs S., Molina M., Blackstock W., Gil C.,
 RA Nombela C.;
 RT "A proteomic approach for the study of Saccharomyces cerevisiae cell
 wall biogenesis.";
 RL Electrophoresis 21:3396-3410 (2000).
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor

Qy	3	KILP-	IITLFPSSVLFPAVSADKIPGDESI---	TNIEGPDRDRESSPKINI---	LNNH	54
D6	8	KVLFGI	VILLMAGVFAFSCNV----	GGSSICAGTTSVVNLDPVIQPCONLVLDLSQH	63	
Qy	55	ITAYSESHLYDRMTFLCLSSH-NTLNGACPTSENPPSSSVSGETNITLOFTFKESLIKR	113			
D6	64	ISCNDYCGWYD-----TDHINLVQG-----SAPAG-----	89			
Qy	114	ELOQTGKY-QLLPFKSVCPSGLTNSAHPNCNKNAASGASLYXIYPAGELKNLPFGGIWD	172			
D6	90	--SLQSQYKGSYLWNKVVPFFLTNTNVLDIGDKTPMLPLPLYI-----TPVGAAGG	140			
Qy	173	ATLK-----LRVKARYSTYGT-----YTINIIIR-----LTDKGNIOIWLPQPK	212			
D6	141	VVIRAGEVIARIHMVKIATLGSGNPRTFNWIIISNNVMVMPFGGCTVDSRNVTVDLDPDP	200			
Qy	213	SDARVDLNRPTGGTYTGIRNSVDMCFDYGSTNSSLSLEIRFDQNPXKSDGKFYLARKIND	272			
D6	201	GSAIPL-----GVYCS-SBQKLUSFYLSGAIOTSSRQV-FANTAP-----D	239			
Qy	273	DTKEIAVTLSSLACK-----SLPTNGTSLNIADAASLETWNRIATAVIMPEI	321			
D6	240	NTKASCVAVTMENCKIATCNVSLGTGVNKKSVPLGISATYGCOTGNYSAGIVSQSV	296			

-!- SIMILARITY: Belongs to the phenylalanyl-tRNA synthetase beta chain family. Subfamily 1.

-!- SIMILARITY: Contains 1 tRNA-binding domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AL596167; CAC96416.1; -.
PIR; AH1580; AH1580.
ListIDst; LIN01648; -.
HAMAP; MF_00283; ; 1.
InterPro; IPR005146; B3_4.
InterPro; IPR005147; B5.
InterPro; IPR005121; Fdx-AnticB.
InterPro; IPR008994; Nucleic acid_OB.
InterPro; IPR004532; PheI bact.
InterPro; IPR002547; trNA_bind.
Pfam; PF03483; B3_4; 1.
Pfam; PF03484; B5; 1.
Pfam; PF03447; FDX-ACS; 1.
Pfam; PF01588; trNA_bind; 1.
TIGRFAMs; TIGR00472; pheI_bact; 1.
PROSITE; PS50886; TRBD; 1.
Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
Metal-binding; Magnesium; RNA-binding; trNA-binding;
Complete proteome.
DOMAIN 40 155 trNA-BINDING.
METAL 462 462 MAGNESIUM (BY SIMILARITY).
METAL 468 468 MAGNESIUM (VIA CARBOXYL OXYGEN) (BY
SIMILARITY)
METAL 471 471 MAGNESIUM (BY SIMILARITY).
METAL 472 472 MAGNESIUM (BY SIMILARITY).
SEQUENCE 802 AA; 88191 MW; D75802193964D7D0 CRC64;

Query Match 5.2%; Score 99; DB 1; Length 802;
Best Local Similarity 17.3%; Pred.No.6.7; Indels 140; Gaps 15;
Matches 70; Conservative

QY 48 HNILNNHITAYSSHTLYDYRMFTICLSHSNPLNGACPTSNPFSSSVSGETNLTQFTEK 107
DB 422 NRILGTSL-SUSESETIFDRLGFLVKEDTLITEVPTR-----RWDTITIE---- 466

QY 108 RSLIKRELQIKGYKQLLFKSVCNPSGLTLNSAHFNCKNAASGASLIYLIPAGELKNLPF 167
DB 467 ADILEEVARIYGDEI---PVTLPATSTV-----TIKLTKGNIOIWLPOFKSDRV 492

QY 168 GGIMDAATLKVRKRYSSTYG-----TYTKN-----TIKLTKGNIOIWLPOFKSDRV 217

DB 493 GGLSDSQKARKVMRAYLEGCAGNQALTISLSKKDATRLAISDEKTVALSMPESESHL 552

QY 218 DLNLRP---TGCGTVIGNSWVMCFYD---GYSTNSSSLERFQ----- 255

DB 553 RTSIVFQLIRASNYNIARKMDVALYENGTFVFYATEGDNLPIEQEHLAGLTGNWHTADW 612

QY 256 DNNPKS-----DGKFYLRKINDDKETIAYTLLSIAGK-----SL 290

DB 613 QKTPKPDFVVLKGIVEGLVNKLGI EAELHWQIKESLHPQRTASI QLEGKEI GYLGA 672

QY 291 TPNTGTSINADI AASLETNWNEITA VTMPEISV-PVLCWE----- 329

DB 673 HPAVEASYDLKETYYVEINV KALLDATKEKVYHPHPRYPETMRDIALLVKD TDHATIS 732

QY 330 -----GRLOLDKAVNPENPCGYMGN-----INVTFPPSSQTLL 361

DB 733 QUIKEHGNNLLVDIELFDIFESESIGENKKS LAYILTFFLDSERTL 777

Db 862 KE-NSHWELTGN---NVNQLMLTNG---HHHLNAQDAN-----KV 896

Qy 186 ETYGTNTNITIKLTKDNKIQLWLPQKSDA-RVDLNLRLPTGGTY-----IGR-NSVDM 237

Db 897 TYNTLTWN---SLSGNSFYVWDFTNKSKVNVKSAIGNFLQVADTKGEHNL 953

Qy 238 CFYDGYSTNSSLEIRFODNN-PSKDGKPYLRKINDTKIEIATYLSLLAGKSLFTPTNGT 296

Db 954 TLFDSATERNLEVLTVANGSVDRGAWKYLRVNG--RYDLYNPEVEKRNQTVJTNIT 1011

Qy 297 SUN--LADAASLETNWRITATMP 319

Db 1012 TPNDIQADAPSAQSNNEIARVETP 1036

RESULT 10

ID YD93 METJA STANDARD; PRT; 608 AA.

AC Q58788;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein MJ1393.

GN MJ1393.

OS Methanococcus jannaschii.

OC Archaea; Euryarchaeota; Methanococci; Methanococcales;

OC Methanocaldococcaceae; Methanocaldococcus.

OX NCBI_TaxID=2190;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;

RX MEDLINE=96337999; PubMed=8688087;

RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,

RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,

RA Kervase A.R., Dougherty B.A., Tomb J.P., Adams M.D., Reich C.L.,

RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Gloeck A.,

RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,

RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,

RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,

RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;

RT "Complete genome sequence of the methanogenic archaeon, Methanococcus

RT jannaschii."

RL Science 273:1058-1073 (1996).

CC -1- SIMILARITY: TO M.JANNASCHII MJ1394 AND A.FULGIDUS AF2028.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

CC EMBL; U67579; AAB99403.1; -

CC PIR; H64473; H64473.

CC TIGR; MJ1393; -

KW Hypothetical protein; Transmembrane; Complete proteome.

FT TRANSSEQ 4 24 POTENTIAL.

SQ SEQUENCE 608 AA; 66768 MW; 010FAP1C29F9C73C CRC64;

Query Match

Best Local Similarity 20.3%; Pred.No.5.1;

Matches 85; Conservative 68; Mismatches 119; Indels 147; Gaps 25;

Qy 1 MNKILF---IPTLPSSVLFTFAVSADKIP-----GDESTNIFGPRDR 41

Db 1 MRKLIFALLMSLIFGTFVGY---GDNGPLVAYVEXYNITGNITGGLVSSIT----- 52

Qy 42 NESSPKINLNHITAYSESHLYDRMTFLCLSSHNT-----LNG-----ACPTSE 87

Db 53 -ESTIGYIVINN--TGTTINDTLVDVVAVNISNNITGPEVYVNGTPKGFIESAPAYT 109

Qy 88 N-PSSSS-----VSGETNITLQTEKSLIKRELQ-KGYKQLLFKSVNCPSGLTLN-SA 139

Db 110 NLPVNTYIHIPIPNNSYVIKFAIDKSIITGVPLINE-----TYSDTKIFSERLSNWSV 165

Qy 140 HFNCKNAASASLYLYIPAGE-----LKNLPFGGIWDATLKLVRKRISETYGT 189

Db 166 YLNIISRNVA-----LPATDTPVSVIMTKYLSNDP-----NNYGS 200

Qy 190 YT---INITIKLTKGNTQIW-----LPQFK-----SDARVDLNLRLPTGGT 228

Db 201 DTWPFNTGTAIANGESTILWDGPFYLFEGYDLSLTWTGVTNNKAITNII--TGNT 258

Qy 229 YIGRNSVDWCFYDGYSTNSSLEIRFODNNPKSDGKPYLRKINDTKIEIATP-----L 283

Db 259 YNRTGTLMKY--GPV-----TPFENGYSKTKI-----EGIVATGYGVSA 300

Qy 284 LIAGKSLPTNG-----TSLNIADAASLETNWRITATMPETI--SVPVLCPORLQL 334

Db 301 TKEGPFNLASSGKIETWESANVSKAS--SYFNLTHVTIWAANGSFVILDPFNITL 357

RESULT 11

ID SP21 YEAST STANDARD; PRT; 758 AA.

AC P35209;

DT 01-FEB-1994 (Rel. 26, Created)

DT 01-FEB-1994 (Rel. 26, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE SPT21 protein.

GN SPT21 OR YMR179W OR YN8010.09.

OS Saccharomyces cerevisiae (Baker's Yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RC MEDLINE=94186069; PubMed=8138180;

RX Natsoulis G., Winston F., Boeke J.D.;

RA "The SP110 and SPT21 genes of Saccharomyces cerevisiae."

RT Genetics 136:93-105 (1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=S288C / AB972;

RX MEDLINE=97313268; PubMed=9169872;

RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,

RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,

RA Jagels K., Lye G., Moule S., Odeil C., Pearson D., Rajandream M.A.,

RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;

RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome

RT XII."

RL Nature 387:90-93 (1997).

CC -1- FUNCTION: Required for normal transcription at a number of loci in

CC yeast.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

CC EMBL; L24436; AAA5078.1; -

CC EMBL; Z49808; CAA89912.1; -

CC PIR; S47866; S47866.

CC GenOnline; 142853; -

CC TRANSFAC; T04376; -

CC SGD; S0004751; SPT21.

CC GC; G0:0006357; P:regulation of transcription from Pol II pro. . .; IMP.

CC DOMAIN 127 144 ASP/GLU-RICH (ACIDIC).

CC FT DOMAIN 672 682 ASP/GLU-RICH (ACIDIC).

SQ SEQUENCE 758 AA; 84697 MW; 7DB3FCF7EE996705 CRC64;

Query Match

5.2%; Score 98; DB 1; Length 758;

GO; GO:006406; P-mRNA-nucleus export; IMP.
GO; GO:006999; P:nuclear pore organization and biogenesis; IMP.
DR InterPro; IPR007187; NupI33.
DR Pfam; PF04044; NupI33; 1.
KW Nuclear protein; Transport.
SQ SEQUENCE 1157 AA; 13319 MW; C8BDBB7D709C5C8 CRC64;

Query Match 5.2%; Score 98; DB 1; Length 1157;
Best local similarity 22.1%; Pred.No.13;
Matches 75; Conservative 43; Mismatches 127; Indels 94; Gaps 21;

QY 43 HNTLNHHITAY-----SSSHLYIDRMTF-----LCLSHNTLNGACPTSENP 89
 :
Db 96 YGVVNDHKKVIWNHSTOKDTPIIVPFRSDNDEIAVAPRCLITFPATMDESPLALNP 155
 :
QY 90 SSSSVSGE-----TNTLTQTEKRSLIKRELQIK-----CYKQLLFK 126
 :
Db 155 NDQDETGGLLIKGSKAIYYEDINSINNLFKLSEKFS---HELELPINSSGGEKCDLM- 211
 :
QY 127 SVNC-SGLTLNS-----AHPCNKNAASGSLVLIYPAGELKNLPG-GIWDA TLKL R-- 178
 :
Db 212 -LACEPAGIVLSTMGRIFFITTIRSMGKSQLKL----GKLLNKKPKLGHWISKFNTNSS 266
 :
QY 179 -VKRRYSETY--GYTINIITIKLTGKNIQIWLPOFQS DARVDLNL RFTGGGTIGNSV 235
 :
Db 267 VSLRNGLPIGKTRLVIIIT---TNKGIFQTW-QLSA-----TNSHPT-----KLI 308
 :
QY 235 DMCFYDCYSTNSSSL-----EIFQQNNPKSD--GRFYLEKINDPT-KSIATVLSLLJA 286
 :
Db 309 DVNIYEAILSELQDYFFAFGLTKIWDSPHLQDESSQLFLSSIYDSCNETYVLSTIIF 368
 :
QY 287 GKSLTPNGTSLNIADAASLETWKNRITAVTM-PEISVP 324
 :
Db 369 DSS-----SNSITFSTRINTFMESITDTKFKPKTIIP 402
 :

RESULT 13
CID2_PAEMA
ID CDG2 PAEMA STANDARD; PRI; 713 AA.
AC P31835;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, last sequence update)
DT 10-OCT-2003 (Rel. 42, last annotation update)
DE Cyclomalto-dextrin glucanotransferase precursor (EC 2.4.1.19)
DE (Cyclodextrin-glycosyltransferase) (CGTase).
OS Paenibacillus macerans (Bacillus macerans).
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.
RX NCBI_Taxid=44252;
RN [1]
SEQUENCE FROM N.A., AND SEQUENCE OF 28-37.
RA Sugimoto T., Kubota M., Sakai S.;
RT "polypeptide possessing cyclomalto-dextrin glucanotransferase activity";
RL Patent number GB2169902, 23-JUL-1986.
CC -!- CATALYTIC ACTIVITY: Degrades starch to cyclodextrins by formation of a 1,4-alpha-D-glucosidic bond.
CC -!- COFACTOR: Binds 2 calcium ions per subunit (By similarity).
CC -!- SUPUNIT: Monomer.
CC -!- MISCELLANEOUS: CGTASE MAY CONSIST OF TWO PROTEIN DOMAINS: THE ONE IN THE AMINO-TERMINAL SIDE CLEAVES THE ALPHA-1,4-GLUCOSIDIC BOND IN STARCH, AND THE OTHER IN THE C-TERMINAL SIDE CATALYZES OTHER ACTIVITIES, INCLUDING THE RECONSTITUTION OF AN ALPHA-1,4-GLUCOSIDIC LINKAGE FOR CYCLIZING THE MALTOOLIGOSACCHARIDE PRODUCED.
CC -!- SIMILARITY: Belongs to family 13 of glycosyl hydrolases.
DR PIR; S26589; ALBSXR.
DR HSSP; P43379; LCDG.
DR InterPro; IPR006589; Alp_aryl_cat_sub.
DR InterPro; IPR006048; Alpha_aryl_C.
DR InterPro; IPR006047; Alpha_aryl_cat.
DR InterPro; IPR002044; CSD_4.
DR InterPro; IPR006046; Glyco_hydro_13.
DR InterPro; IPR007110; Ig-like.

	Best Local Similarity	21.9%;	Pred. No.	7.4;	Matches	61; Conservative	34; Mismatches	95; Indels	88; Gaps	12;
QY	42	NESPKNILNHHITAYSSHTLYDRMTF-----LCLSSHNTLCACPTSPNPSSSVS	95	:	:	:	:	:	:	:
DB	166	NISMKGRVNNQI----PEETLEVLRFTKVIUNRTSGNNTTNSRSLCLOMPSSL---	218	:	:	:	:	:	:	:
QY	96	GETHNITLFQTEKRSLIKRELIQKGFKQLAFKVCNCPGSLTMSAHFNCKNAASGLYL	155	:	:	:	:	:	:	:
DB	219	--PRAATLPFTPKSLSLKTNQIK-----NSRNAATTITIN-----NINSCT---	257	:	:	:	:	:	:	:
QY	156	YIPAGELKNLPFGGIWDATJLVKVRYSYTYGTYNITIKLTDKGNIQIWPOFKSDA	215	:	:	:	:	:	:	:
DB	258	--VGRRQTNEPM---PAFXAVRTQ-----SUPW-----	280	:	:	:	:	:	:	:
QY	216	RVDNLNRPTGGGVIGNSVDMCFY-----DGYSTNSSLEIRFDONNPKDGKFYLRK	269	:	:	:	:	:	:	:
DB	281	----NLKENIANTFPPNSIAHKYYLADRTEANQQNKQHONTAYEINTLONDNTIQRTK	336	:	:	:	:	:	:	:
QY	270	INDD--TKBIAYTLILLLAGKSLLTPNGT-----SLNI	300	:	:	:	:	:	:	:
DB	337	IDDSVSXRFDFWMKRXSKTKVSGFIATIAKKPASINI	374	:	:	:	:	:	:	:
<hr/>										
RESULT 12										
N133 YEAST										
ID	N133 YEAST	STANDARD;	PRT;	1157 AA.						
AC	P36161;									
DT	01-JUN-1994 (Rel. 29, Created)									
DT	01-JUN-1994 (Rel. 29, Last sequence update)									
DT	15-MAR-2004 (Rel. 43, Last annotation update)									
DE	Nuc.eoporin NUP133 [Nuclear pore protein NUP133].									
DE	NUP133 OR YKR082W OR YXR402.									
GN	Saccharomyces cerevisiae (Baker's yeast).									
OS	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;									
OCC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.									
OC	NCBI_TAXID=4932;									
[1]	SEQUENCE FROM N.A.									
RP	STRAIN=JUOXR;									
RC	MEDLINE=95112817; PubMed=7813444;									
RX	Doye V., Wepf R., Hurt B.C.;									
RA	"A novel nuclear pore protein Nup133p with distinct roles in poly(A)+									
RT	RNA transport and nuclear pore distribution.";									
RL	EMBO J. 13:6062-6075(1994).									
RN	[2]									
SN	SEQUENCE FROM N.A.									
XX	MEDLINE=94362327; PubMed=8203164;									
XA	Garcia-Cantalejo J., Baladron V., Esteban P.F., Santos M.A., Bou G.,									
RA	Remacha M.A., Revuelta J.L., Ballista J.P.G., Jimenez A., del Rey F.;									
RT	"The complete sequence of an 18,002 bp segment of Saccharomyces									
RT	cerevisiae chromosome XI contains the HBS1, MRP-L20 and PRP16 genes,									
RT	and six new open reading frames.";									
RL	Yeast 10:231-245(1994).									
-!	FUNCTION: Involved in poly(A) + RNA transport and nuclear pore									
CC	distribution.									
-!	SUBCELLULAR LOCATION: Nuclear pore complex.									
<hr/>										
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce or send an email to licensed@isb-sib.ch).										
CC	EMBL; X80066; CAA36372.1 ; -									
DR	EMBL; Z27116; CAA81633.1 ; -									
DR	EMBL; Z28307; CAA82161.1 ; -									
DR	PIR; S38160; S38160.									
DR	GermOnline; 140061; -									
DR	SED; SC000790; NUP133.									
DR	GO; GO:0005643; C:nuclear pore; IDA.									

```
DR InterPro; IPR002909; IPT TIG.
DR Pfam; PF00128; alpha-amylase; 1.
DR Pfam; PF02806; alpha-amylase_C; 1.
DR Pfam; PF02886; CBM_20; 1.
DR Pfam; PF01833; TIG; 1.
DR PRINTS; PRO1110; ALPHAAMYLASE.
DR ProDom; PD001568; CBD 4; 1.
DR SMART; SM00632; AmyC; 1.
DR Transferase; Glycosyltransferase; Calcium-binding; Signal.
FT SIGNAL 1 27
FT CHAIN 28 713
FT DOMAIN 28 165 A1.
FT DOMAIN 166 229 B.
FT DOMAIN 230 434 A2.
FT DOMAIN 435 522 C.
FT DOMAIN 523 609 D.
FT DOMAIN 610 713 E.
FT ACT_SITE 256 256
FT ACT_SITE 285 285
FT ACT_SITE 356 356
FT METAL 54 54
FT METAL 56 56
FT METAL 59 59
FT METAL 60 60
FT METAL 78 78
FT METAL 80 80
FT METAL 156 156
FT METAL 217 217
FT METAL 226 226
FT METAL 260 260
SQ SEQUENCE 713 AA; 76857 MW; 5A287BCC4AAFE635 CRC64;

Query Match 5.1%; Score 95.5; DB 1; Length 713;
Best Local Similarity 20.1%; Pred. No. 11;
Matches 65; Conservative 43; Mismatches 131; Indels 85; Gaps 12;

QY 68 MTFICLSHNTLNGACPTSENPPSSSVSGETNITLQTEKRSLKRLQKGYKQLLFKS 127
D 8 LTVSVMSVGIAGALPWAAP-DTSVNNKLNFS---TDTVYQIVTDRFVDG-----NS 57
QY 128 VNCPSGHTLSAHNCKNAASGASLYLYIPAGELXNLPFGGIWDATLKLVRKERYSETY 187
D 58 ANNPFGAFAFSDHGN-----LKY-----FGDWDG-----ITNKINDGY 92
QY 188 GTYVITNITIKLTKGTQIWLPPQFKSDARVDLNLRTGGTYIG----- 231
D 93 -----LTGNGITALWISQPVENITAVINYSVGNNTAYHGYWPRDKKTNAAFGSF 142
QY 232 -----RNSVDMCFYDGVSTNSLSLEIRFQDNPKSDGKFVLRKINDGTKEIA 278
D 143 TDFGNLIAAAHSHNIKVVYDPAPNHTNPASSTDFSAENGALYNNGTLLGYSNDTGLF 202
QY 279 YTLISLLAGKSLTPT-NGTSTNIADAASLETNWRNRTAVTMTPEISVPVLCWPGRLQDAK 337
D 203 HH-----NGGTDFSTTESGIYKNLYDLADINQNNNTIDSYLKESQLWNLGVGIRFDV 258
QY 338 VENPRAGQ--YMGNT-----NVITPT 355
D 259 KEMPQGWQKSVSSSIYSSANPVET 282

RESULT 14
REFIM YEAST
AC P30775;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

Peptide chain release factor 1, mitochondrial precursor (MRF-1).
MRF1 OR YGLI43C.
Saccharomyces cerevisiae (Baker's yeast).
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycos.
NCBI_TaxID=4932;
[1]
SEQUENCE FROM N.A. PubMed=1475194;
MEDLINE=93117110;
Pel H.J., Maat M.J., Rep M., Grivell L.A.;
"The yeast nuclear gene MRF1 encodes a mitochondrial peptide chain
release factor and cures several mitochondrial RNA splicing
defects.";
Nucleic Acids Res. 20:6339-6346(1992).
[2]
SEQUENCE FROM N.A.
STRAIN=S288C / FY1769;
MEDLINE=97197983; PubMed=9046099;
Voet M., Defoor E., Verhasselt P., Riles L., Robben J., Velckaert G.;
"The sequence of a nearly unclonable 22.8 kb segment on the left arm
of chromosome VII from Saccharomyces cerevisiae reveals ARO2, RPL5A,
TRP1, MRF1 genes and six new open reading frames.";
Yeast 13:177-182(1997).
-!- FUNCTION: Mitochondrial peptide chain release factor that directs
the termination of translation in response to the peptide chain
termination codons UAA and UAG.
-!- SUBCELLULAR LOCATION: Mitochondrial.
-!- SIMILARITY: Belongs to the prokaryotic/mitochondrial release
factor family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
-----
EMBL; X60381; CAA42932.1; --
EMBL; X99960; CAA68219.1; --
EMBL; Z72665; CAA96855.1; --
PIR; S28602; S28602.
GeneOnline; 141191. --
SGD; S0003111; MRF1.
GO; GO:0005739; C:mitochondrion; IDA. factor activity; IMP.
GO; GO:0003747; F:translation release factor activity; IMP.
GO; GO:0006415; P:translational termination; IMP.
InterPro; IPR005139; PCRF.
InterPro; IPR000352; Pep_rel_factor_1.
Pfam; PF03452; PCRF; 1.
Pfam; PF00472; RF-1; 1.
PROSITE; PS00745; RF_PROK_1; 1.
Protein biosynthesis; Mitochondrion; Transit peptide.
TRANSIT ? ? MITOCHONDRION (POTENTIAL).
CHAIN ? 413 PEPTIDE CHAIN RELEASE FACTOR 1.
SEQUENCE 413 AA; 46770 MW; AAAD4829748C7604 CRC64;

Query Match 5.0%; Score 95; DB 1; Length 413;
Best Local Similarity 21.8%; Pred. No. 5.7;
Matches 67; Conservative 47; Mismatches 121; Indels 72; Gaps 14;

QY 30 SSSSVSGETN--ITLQTEKRSLI-----KELQIKGYKQLLFKSVNCPGSLTNSAHFN 142
D 30 TSTTNSKNGSIPTQYTELSPLLVKQAEKVEALKD-----LDRKLSG-----GIHFD 78
QY 143 CNK-----NAASGASLYLYIPAGELXNLPFGGIWDATLKLVRKERYSE---TYGIYTI 192
D 79 VNKQHYAKLSALTDTT--EYKELKLNELKSLQEMIVSDPSLRAAEAEQYAEALVPPQYETISS 238
QY 193 NITIKLTKGNIQIWLPPQFKSDARVDLNLRTGGG---TYIGRNSVDMCFYDGVSTNSSS 249
D 139 RLNVKLLP-----PEPFADKPSLLELPDGVGGIEAMIFQNLDM--YIGYA-NYRK 187
```

QV 250 LEIRFQDNPKSGDKFYLRKNDTKEIAYTSLLLACK-----SITPTNGTSLNTADA 303
Dd 188 WKYALISKVENESGGIADALLSIEAGSYDRLAFEGVHRVQIPSTETRG----- 239
QV 304 ASLTNNWRITAVTMPEI-----SVPLVCPGRQLDAKVENPEAGQVMGN---I 350
Dd 240 ---RTHSTAATAVVVLPQIGDSESAKSIDAYETFRFGEIRVDIMRASGGGGRVNTDSAV 296
QV 351 NVITPES 357
Dd 297 RLTHIPS 303
RESULT 15
WAPA_BACSU
ID WAPA_BACSU STANDARD; PRT; 2334 AA.
AC Q07833;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Wall-associated protein precursor.
GN WAPA OR N17G OR BSU39230.
OS Bacillus subtilis.
OS Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=93302506; PubMed=8316082;
RA Foster S.J.;
RT Molecular analysis of three major wall-associated proteins of
RT Bacillus subtilis 168: evidence for processing of the product of a
RT gene encoding a 258 kDa precursor two-domain ligand-binding
RT Protein."
RL Mol. Microbiol. 8:299-310(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSC1A1;
RX MEDLINE=95219088; PubMed=7704263;
RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
RT genome containing the hut and wapa loci."
RL Microbiology 141:337-343(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSC1A1;
RX MEDLINE=97124196; PubMed=8969509;
RA Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
RA Miwa Y., Fujita Y.;
RT "Sequencing of a 65 kb region of the Bacillus subtilis genome
RT containing the lic and cel loci, and creation of a 177 kb contig
RT covering the gnt-sacxy region."
RL Microbiology 142:3113-3123(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Besieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Briegleb S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizet F., Devine K.M., Dusterhoft A., Ehrlich S.J., Emmerson P.T.,
RA Etian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Gim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Hailech S., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Ketterer P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Iardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,

RA Noone D., O'Reilly M., Ogiwara K., Oudega A., Oudega B., Park S.H.,
RA Paro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Prassecan E., Pujic P., Purrelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger N., Rivolta C., Rocha E., Roche S., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemart K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpatra P., Tognoni A.,
RA Toseato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
CC -!- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM,
CC MOTILITY, SECRETION OR DIFFERENTIATION.
CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED
CC INTO THE MEDIUM.
CC -!- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS; THE N-TERMINUS, HAS THREE
CC 101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE
CC C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED
CC MOTIF REPEATED 31 TIMES.
CC -!- SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME
CC SIMILARITY TO THE REPEAT IN E.COLI Rhs GROUP OF PROTEINS (RHS-A-D).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by not-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; L05634; AAA22883.1; -
DR EMBL; D31856; BAA06656.1; -
DR EMBL; D29885; BAA06260.1; -
DR EMBL; D83026; BAA11693.1; -
DR EMBL; S29124; CAB15959.1; -
DR FIR; S32920; S32920.
DR Subtilisin; BG10797; wapA.
DR InterPro; IPR003305; CBM_CenC.
DR InterPro; IPR006530; YD_
DR Pfam; PF02018; CBM_4_9; 1.
DR Pfam; PF05893; Rhs_repeat; 14.
DR TIGRfam; TIGR01643; YD_repeat_2x; 17.
KW Cell wall; Repeat; Signal; Complete proteome.
OR 32 (POTENTIAL).
FT SIGNAL 1 28
FT CHAIN 29 2334
FT DOMAIN 504 869
FT REPEAT 504 605
FT REPEAT 636 736
FT REPEAT 769 869
FT DOMAIN 1021 2139
FT REPEAT 1021 1040
FT REPEAT 1042 1061
FT REPEAT 1063 1082
FT REPEAT 1083 1102
FT REPEAT 1109 1128
FT REPEAT 1129 1148
FT REPEAT 1150 1169
FT REPEAT 1174 1193
FT REPEAT 1199 1218
FT REPEAT 1219 1238
FT REPEAT 1238 1257
FT REPEAT 1257 1276
FT REPEAT 1276 1295
FT REPEAT 1295 1314
FT REPEAT 1314 1333
FT REPEAT 1333 1352
FT REPEAT 1352 1371
FT REPEAT 1371 1390
FT REPEAT 1390 1409
FT REPEAT 1409 1428
FT REPEAT 1428 1447
FT REPEAT 1447 1466
FT REPEAT 1466 1485
FT REPEAT 1485 1504
FT REPEAT 1504 1523
FT REPEAT 1523 1542
FT REPEAT 1542 1561
FT REPEAT 1561 1580
FT REPEAT 1580 1599
FT REPEAT 1599 1618
FT REPEAT 1618 1637
FT REPEAT 1637 1656
FT REPEAT 1656 1675
FT REPEAT 1675 1694
FT REPEAT 1694 1713
FT REPEAT 1713 1732
FT REPEAT 1732 1751
FT REPEAT 1751 1770
FT REPEAT 1770 1789
FT REPEAT 1789 1808
FT REPEAT 1808 1827
FT REPEAT 1827 1846
FT REPEAT 1846 1865
FT REPEAT 1865 1884
FT REPEAT 1884 1903
FT REPEAT 1903 1922
FT REPEAT 1922 1941
FT REPEAT 1941 1960
FT REPEAT 1960 1979
FT REPEAT 1979 1998
FT REPEAT 1998 2017
FT REPEAT 2017 2036
FT REPEAT 2036 2055
FT REPEAT 2055 2074
FT REPEAT 2074 2093
FT REPEAT 2093 2112
FT REPEAT 2112 2131
FT REPEAT 2131 2150
FT REPEAT 2150 2169
FT REPEAT 2169 2188
FT REPEAT 2188 2207
FT REPEAT 2207 2226
FT REPEAT 2226 2245
FT REPEAT 2245 2264
FT REPEAT 2264 2283
FT REPEAT 2283 2302
FT REPEAT 2302 2321
FT REPEAT 2321 2340
FT REPEAT 2340 2359
FT REPEAT 2359 2378
FT REPEAT 2378 2397
FT REPEAT 2397 2416
FT REPEAT 2416 2435
FT REPEAT 2435 2454
FT REPEAT 2454 2473
FT REPEAT 2473 2492
FT REPEAT 2492 2511
FT REPEAT 2511 2530
FT REPEAT 2530 2549
FT REPEAT 2549 2568
FT REPEAT 2568 2587
FT REPEAT 2587 2606
FT REPEAT 2606 2625
FT REPEAT 2625 2644
FT REPEAT 2644 2663
FT REPEAT 2663 2682
FT REPEAT 2682 2701
FT REPEAT 2701 2720
FT REPEAT 2720 2739
FT REPEAT 2739 2758
FT REPEAT 2758 2777
FT REPEAT 2777 2796
FT REPEAT 2796 2815
FT REPEAT 2815 2834
FT REPEAT 2834 2853
FT REPEAT 2853 2872
FT REPEAT 2872 2891
FT REPEAT 2891 2910
FT REPEAT 2910 2929
FT REPEAT 2929 2948
FT REPEAT 2948 2967
FT REPEAT 2967 2986
FT REPEAT 2986 3005
FT REPEAT 3005 3024
FT REPEAT 3024 3043
FT REPEAT 3043 3062
FT REPEAT 3062 3081
FT REPEAT 3081 3100
FT REPEAT 3100 3119
FT REPEAT 3119 3138
FT REPEAT 3138 3157
FT REPEAT 3157 3176
FT REPEAT 3176 3195
FT REPEAT 3195 3214
FT REPEAT 3214 3233
FT REPEAT 3233 3252
FT REPEAT 3252 3271
FT REPEAT 3271 3290
FT REPEAT 3290 3309
FT REPEAT 3309 3328
FT REPEAT 3328 3347
FT REPEAT 3347 3366
FT REPEAT 3366 3385
FT REPEAT 3385 3404
FT REPEAT 3404 3423
FT REPEAT 3423 3442
FT REPEAT 3442 3461
FT REPEAT 3461 3480
FT REPEAT 3480 3499
FT REPEAT 3499 3518
FT REPEAT 3518 3537
FT REPEAT 3537 3556
FT REPEAT 3556 3575
FT REPEAT 3575 3594
FT REPEAT 3594 3613
FT REPEAT 3613 3632
FT REPEAT 3632 3651
FT REPEAT 3651 3670
FT REPEAT 3670 3689
FT REPEAT 3689 3708
FT REPEAT 3708 3727
FT REPEAT 3727 3746
FT REPEAT 3746 3765
FT REPEAT 3765 3784
FT REPEAT 3784 3803
FT REPEAT 3803 3822
FT REPEAT 3822 3841
FT REPEAT 3841 3860
FT REPEAT 3860 3879
FT REPEAT 3879 3898
FT REPEAT 3898 3917
FT REPEAT 3917 3936
FT REPEAT 3936 3955
FT REPEAT 3955 3974
FT REPEAT 3974 3993
FT REPEAT 3993 4012
FT REPEAT 4012 4031
FT REPEAT 4031 4050
FT REPEAT 4050 4069
FT REPEAT 4069 4088
FT REPEAT 4088 4107
FT REPEAT 4107 4126
FT REPEAT 4126 4145
FT REPEAT 4145 4164
FT REPEAT 4164 4183
FT REPEAT 4183 4202
FT REPEAT 4202 4221
FT REPEAT 4221 4240
FT REPEAT 4240 4259
FT REPEAT 4259 4278
FT REPEAT 4278 4297
FT REPEAT 4297 4316
FT REPEAT 4316 4335
FT REPEAT 4335 4354
FT REPEAT 4354 4373
FT REPEAT 4373 4392
FT REPEAT 4392 4411
FT REPEAT 4411 4430
FT REPEAT 4430 4449
FT REPEAT 4449 4468
FT REPEAT 4468 4487
FT REPEAT 4487 4506
FT REPEAT 4506 4525
FT REPEAT 4525 4544
FT REPEAT 4544 4563
FT REPEAT 4563 4582
FT REPEAT 4582 4601
FT REPEAT 4601 4620
FT REPEAT 4620 4639
FT REPEAT 4639 4658
FT REPEAT 4658 4677
FT REPEAT 4677 4696
FT REPEAT 4696 4715
FT REPEAT 4715 4734
FT REPEAT 4734 4753
FT REPEAT 4753 4772
FT REPEAT 4772 4791
FT REPEAT 4791 4810
FT REPEAT 4810 4829
FT REPEAT 4829 4848
FT REPEAT 4848 4867
FT REPEAT 4867 4886
FT REPEAT 4886 4905
FT REPEAT 4905 4924
FT REPEAT 4924 4943
FT REPEAT 4943 4962
FT REPEAT 4962 4981
FT REPEAT 4981 5000
FT REPEAT 5000 5019
FT REPEAT 5019 5038
FT REPEAT 5038 5057
FT REPEAT 5057 5076
FT REPEAT 5076 5095
FT REPEAT 5095 5114
FT REPEAT 5114 5133
FT REPEAT 5133 5152
FT REPEAT 5152 5171
FT REPEAT 5171 5190
FT REPEAT 5190 5209
FT REPEAT 5209 5228
FT REPEAT 5228 5247
FT REPEAT 5247 5266
FT REPEAT 5266 5285
FT REPEAT 5285 5304
FT REPEAT 5304 5323
FT REPEAT 5323 5342
FT REPEAT 5342 5361
FT REPEAT 5361 5380
FT REPEAT 5380 5399
FT REPEAT 5399 5418
FT REPEAT 5418 5437
FT REPEAT 5437 5456
FT REPEAT 5456 5475
FT REPEAT 5475 5494
FT REPEAT 5494 5513
FT REPEAT 5513 5532
FT REPEAT 5532 5551
FT REPEAT 5551 5570
FT REPEAT 5570 5589
FT REPEAT 5589 5608
FT REPEAT 5608 5627
FT REPEAT 5627 5646
FT REPEAT 5646 5665
FT REPEAT 5665 5684
FT REPEAT 5684 5703
FT REPEAT 5703 5722
FT REPEAT 5722 5741
FT REPEAT 5741 5760
FT REPEAT 5760 5779
FT REPEAT 5779 5798
FT REPEAT 5798 5817
FT REPEAT 5817 5836
FT REPEAT 5836 5855
FT REPEAT 5855 5874
FT REPEAT 5874 5893
FT REPEAT 5893 5912
FT REPEAT 5912 5931
FT REPEAT 5931 5950
FT REPEAT 5950 5969
FT REPEAT 5969 5988
FT REPEAT 5988 6007
FT REPEAT 6007 6026
FT REPEAT 6026 6045
FT REPEAT 6045 6064
FT REPEAT 6064 6083
FT REPEAT 6083 6102
FT REPEAT 6102 6121
FT REPEAT 6121 6140
FT REPEAT 6140 6159
FT REPEAT 6159 6178
FT REPEAT 6178 6197
FT REPEAT 6197 6216
FT REPEAT 6216 6235
FT REPEAT 6235 6254
FT REPEAT 6254 6273
FT REPEAT 6273 6292
FT REPEAT 6292 6311
FT REPEAT 6311 6330
FT REPEAT 6330 6349
FT REPEAT 6349 6368
FT REPEAT 6368 6387
FT REPEAT 6387 6406
FT REPEAT 6406 6425
FT REPEAT 6425 6444
FT REPEAT 6444 6463
FT REPEAT 6463 6482
FT REPEAT 6482 6501
FT REPEAT 6501 6520
FT REPEAT 6520 6539
FT REPEAT 6539 6558
FT REPEAT 6558 6577
FT REPEAT 6577 6596
FT REPEAT 6596 6615
FT REPEAT 6615 6634
FT REPEAT 6634 6653
FT REPEAT 6653 6672
FT REPEAT 6672 6691
FT REPEAT 6691 6710
FT REPEAT 6710 6729
FT REPEAT 6729 6748
FT REPEAT 6748 6767
FT REPEAT 6767 6786
FT REPEAT 6786 6805
FT REPEAT 6805 6824
FT REPEAT 6824 6843
FT REPEAT 6843 6862
FT REPEAT 6862 6881
FT REPEAT 6881 6900
FT REPEAT 6900 6919
FT REPEAT 6919 6938
FT REPEAT 6938 6957
FT REPEAT 6957 6976
FT REPEAT 6976 6995
FT REPEAT 6995 7014
FT REPEAT 7014 7033
FT REPEAT 7033 7052
FT REPEAT 7052 7071
FT REPEAT 7071 7090
FT REPEAT 7090 7109
FT REPEAT 7109 7128
FT REPEAT 7128 7147
FT REPEAT 7147 7166
FT REPEAT 7166 7185
FT REPEAT 7185 7204
FT REPEAT 7204 7223
FT REPEAT 7223 7242
FT REPEAT 7242 7261
FT REPEAT 7261 7280
FT REPEAT 7280 7299
FT REPEAT 7299 7318
FT REPEAT 7318 7337
FT REPEAT 7337 7356
FT REPEAT 7356 7375
FT REPEAT 7375 7394
FT REPEAT 7394 7413
FT REPEAT 7413 7432
FT REPEAT 7432 7451
FT REPEAT 7451 7470
FT REPEAT 7470 7489
FT REPEAT 7489 7508
FT REPEAT 7508 7527
FT REPEAT 7527 7546
FT REPEAT 7546 7565
FT REPEAT 7565 7584
FT REPEAT 7584 7603
FT REPEAT 7603 7622
FT REPEAT 7622 7641
FT REPEAT 7641 7660
FT REPEAT 7660 7679
FT REPEAT 7679 7698
FT REPEAT 7698 7717
FT REPEAT 7717 7736
FT REPEAT 7736 7755
FT REPEAT 7755 7774
FT REPEAT 7774 7793
FT REPEAT 7793 7812
FT REPEAT 7812 7831
FT REPEAT 7831 7850
FT REPEAT 7850 7869
FT REPEAT 7869 7888
FT REPEAT 7888 7907
FT REPEAT 7907 7926
FT REPEAT 7926 7945
FT REPEAT 7945 7964
FT REPEAT 7964 7983
FT REPEAT 7983 8002
FT REPEAT 8002 8021
FT REPEAT 8021 8040
FT REPEAT 8040 8059
FT REPEAT 8059 8078
FT REPEAT 8078 8097
FT REPEAT 8097 8116
FT REPEAT 8116 8135
FT REPEAT 8135 8154
FT REPEAT 8154 8173
FT REPEAT 8173 8192
FT REPEAT 8192 8211
FT REPEAT 8211 8230
FT REPEAT 8230 8249
FT REPEAT 8249 8268
FT REPEAT 8268 8287
FT REPEAT 8287 8306
FT REPEAT 8306 8325
FT REPEAT 8325 8344
FT REPEAT 8344 8363
FT REPEAT 8363 8382
FT REPEAT 8382 8401
FT REPEAT 8401 8420
FT REPEAT 8420 8439
FT REPEAT 8439 8458
FT REPEAT 8458 8477
FT REPEAT 8477 8496
FT REPEAT 8496 8515
FT REPEAT 8515 8534
FT REPEAT 8534 8553
FT REPEAT 8553 8572
FT REPEAT 8572 8591
FT REPEAT 8591 8610
FT REPEAT 8610 8629
FT REPEAT 8629 8648
FT REPEAT 8648 8667
FT REPEAT 8667 8686
FT REPEAT 8686 8705
FT REPEAT 8705 8724
FT REPEAT 8724 8743
FT REPEAT 8743 8762
FT REPEAT 8762 8781
FT REPEAT 8781 8800
FT REPEAT 8800 8819
FT REPEAT 8819 8838
FT REPEAT 8838 8857
FT REPEAT 8857 8876
FT REPEAT 8876 8895
FT REPEAT 8895 8914
FT REPEAT 8914 8933
FT REPEAT 8933 8952
FT REPEAT 8952 8971
FT REPEAT 8971 8990
FT REPEAT 8990 9009
FT REPEAT 9009 9028
FT REPEAT 9028 9047
FT REPEAT 9047 9066
FT REPEAT 9066 9085
FT REPEAT 9085 9104
FT REPEAT 9104 9123
FT REPEAT 9123 9142
FT REPEAT 9142 9161
FT REPEAT 9161 9180
FT REPEAT 9180 9199
FT REPEAT 9199 9218
FT REPEAT 9218 9237
FT REPEAT 9237 9256
FT REPEAT 9256 9275
FT REPEAT 9275 9294
FT REPEAT 9294 9313
FT REPEAT 9313 9332
FT REPEAT 9332 9351
FT REPEAT 9351 9370
FT REPEAT 9370 9389
FT REPEAT 9389 9408
FT REPEAT 9408 9427
FT REPEAT 9427 9446
FT REPEAT 9446 9465
FT REPEAT 9465 9484
FT REPEAT 9484 9503
FT REPEAT 9503 9522
FT REPEAT 9522 9541
FT REPEAT 9541 9560
FT REPEAT 9560 9579
FT REPEAT 9579 9598
FT REPEAT 9598 9617
FT REPEAT 9617 9636
FT REPEAT 9636 9655
FT REPEAT 9655 9674
FT REPEAT 9674 9693
FT REPEAT 9693 9712
FT REPEAT 9712 9731
FT REPEAT 9731 9750
FT REPEAT 9750 9769
FT REPEAT 9769 9788
FT REPEAT 9788 9807
FT REPEAT 9807 9826
FT REPEAT 9826 9845
FT REPEAT 9845 9864
FT REPEAT 9864 9883
FT REPEAT 9883 9902
FT REPEAT 9902 9921
FT REPEAT 9921 9940
FT REPEAT 9940 9959
FT REPEAT 9959 9978
FT REPEAT 9978 9997
FT REPEAT 9997 10000

Tue May 11 10:31:29 2004

FT	REPEAT	1840	1859	2-19.
FT	REPEAT	1861	1880	2-20.
FT	REPEAT	1867	1906	2-21.
FT	REPEAT	1908	1927	2-22.
FT	REPEAT	1929	1948	2-23.
FT	REPEAT	1969	1982	2-24 (APPROXIMATE) .
FT	REPEAT	1983	2002	2-25.
FT	REPEAT	2008	2027	2-26.
FT	REPEAT	2028	2047	2-27.
FT	REPEAT	2051	2070	2-28.
FT	REPEAT	2071	2090	2-29.
FT	REPEAT	2093	2112	2-30.
FT	REPEAT	2120	2139	2-31.
SQ	SEQUENCE	2334	AA; 258329	NW; B75138CCD278BA3 CRC64;

Query Match	5.04; Score 95; DB 1; Length 2334;
Best local Similarity	22.14; Pred. No. 53;
Matches	71; Conservative 42; Mismatches 106; Indels 102; Gaps 20;

QY	29	DESITNI	FGPRDRNRSSPMHILNN	-----HITA-----YSESHLY-----DRMTFLCLS	74
DB	260	DSKLDEL	SGEVERSD	-KVSYKLEKNEEGYL-HLTADENWLKDPREVYPVSDIPSTLSVS	318
QY	75	SHNTLNG	ACPTSENPSSSSNGFENITL	QFTEKRSLIKRELIQIKGYKQLLPKSVNCPSGL	134
DB	319	SDTFVNG	AYPTT-NYSASSQKWDAN	-----LKYA-----	346
QY	135	TLNSAHF	CNCKNAASGASLYLIYI	PAGELKNLPFGGIWDATILKLRVKRRYSYTYGTITINI	194
DB	347	VLKTTY	---DKTGTG-N	YAPMKFNLLKP-QNMVTVKAT-KTVVAHSY---YGT----	393
QY	195	TIKLT	DKNIQIWLDPQFS	---DAEVDNLNRPCTGGGTVIGNSVDMCVDGY-----S	244
DB	394	--KATG	-----LWLTJTVNSYDN	AKVTWNTKPSKN--IGKADVHKGWASVDVTA VKS	444
QY	245	TNSSL	LEIRFOODNPKSDCKPYLRKI	-----NDDTKEIATVLSILLAGKSLTPT----	293
DB	445	WNSGG	ANYGFK-LHNTNGKYE	WKYKLLSSANSANKPYIEVTYITP-----XGNTPTIKAY	498
QY	294	-NGTSL	NIADAASLETTNNRI	313	
DB	499	HNGD	STGYFD-----ISWKV	514	

Search completed: May 6, 2004, 10:18:12
Job time : 20 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

QM protein - protein search, using sw model

Run on: May 6, 2004, 10:16:37 ; Search time 23 Seconds
(without alignments)
810.303 Million cell updates/sec

Title: US-09-839-894-10
Perfect score: 1886
Sequence: 1 MNKILFIFLTFSSVLTFPA.....EAGQYMGKINVTFTSSQTL 361

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/2/iaa/5A COMB.pep.*
2: /cgn2_6/prodata/2/iaa/5B COMB.pep.*
3: /cgn2_6/prodata/2/iaa/6A COMB.pep.*
4: /cgn2_6/prodata/2/iaa/6B COMB.pep.*
5: /cgn2_6/prodata/2/iaa/6CTUS COMB.pep.*
6: /cgn2_6/prodata/2/iaa/6ackfiles1.pep.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1536.5	81.5	360	2	US-08-483-101-17
2	988.5	52.4	363	2	US-08-483-101-16
3	927.5	49.2	364	2	US-08-483-101-5
4	116.5	6.2	1600	2	US-08-617-697-10
5	114	6.0	650	3	US-08-362-525-2
6	111.5	5.9	1035	4	US-09-206-942-41
7	111.5	5.9	1011	4	US-09-206-942-39
8	111.5	5.9	1529	2	US-08-728-470-10
9	111.5	5.9	1529	3	US-08-719-641-10
10	108	5.7	524	4	US-09-242-913B-15
11	107	5.7	1095	4	US-09-206-942-45
12	107	5.7	1101	4	US-09-206-942-43
13	106	5.6	671	2	US-08-737-716-13
14	101	5.4	915	4	US-09-206-942-35
15	101	5.4	1222	4	US-09-206-942-37
16	101	5.4	1228	4	US-09-206-942-34
17	101	5.4	1338	2	US-08-728-470-9
18	101	5.4	1338	3	US-08-719-641-9
19	101	5.4	1599	2	US-08-617-697-9
20	100.5	5.3	1004	4	US-09-206-942-57
21	100.5	5.3	1010	4	US-09-206-942-55
22	99.5	5.3	969	4	US-09-206-942-32
23	99.5	5.3	975	4	US-09-206-942-30
24	99	5.2	992	4	US-09-206-942-61
25	99	5.2	998	4	US-09-206-942-59
26	99	5.2	1848	3	US-08-236-791-6
27	99	5.2	1848	4	US-08-839-996-6

28	59	5.2	1848	4	US-10-080-505-6	Sequence 6, Appli
29	59	5.2	1848	5	PCT-US95-10661A-6	Sequence 6, Appli
30	97	5.1	977	4	US-09-206-942-53	Sequence 53, Appl
31	97	5.1	983	4	US-09-206-942-51	Sequence 51, Appl
32	94.5	5.0	843	4	US-09-543-681A-6880	Sequence 6880, Ap
33	94.5	5.0	1912	1	US-08-409-995-4	Sequence 4, Appli
34	94.5	5.0	1912	3	US-08-685-457-4	Sequence 4, Appli
35	94.5	5.0	2353	3	US-09-377-155-33	Sequence 33, Appl
36	94.5	5.0	2353	3	US-08-911-942-4	Sequence 4, Appli
37	94.5	5.0	2353	4	US-09-669-974-33	Sequence 33, Appl
38	94.5	5.0	2353	4	US-09-797-862-33	Sequence 33, Appl
39	94.5	5.0	2354	4	US-09-268-347-47	Sequence 47, Appl
40	94.5	5.0	2411	4	US-09-268-347-36	Sequence 36, Appl
41	93.5	5.0	715	4	US-09-462-917A-134	Sequence 134, App
42	93.5	5.0	901	4	US-09-134-001C-5351	Sequence 5351, Ap
43	93	4.9	1167	1	US-08-485-568A-6	Sequence 6, Appli
44	93	4.9	1167	2	US-08-590-554A-6	Sequence 6, Appli
45	93	4.9	1167	2	US-09-184-223-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-08-483-101-17
; Sequence 17, Application US/08483101
; Patent No. 5932715
; GENERAL INFORMATION:
; APPLICANT: Scott, June R.
; APPLICANT: Froehlich, Barbara
; APPLICANT: Caron, Judy
; TITLE OF INVENTION: CS2 Proteins and Coding Sequences
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee and Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,101
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Feiber, Donna M.
; REGISTRATION NUMBER: 33878
; REFERENCE/DOCKET NUMBER: 6-95
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 360 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
US-08-483-101-17

Query Match 81.5%; Score 1536.5; DB 2; Length 360;

Best Local Similarity 50.3%; Pred. No. 1.3e-149;

Matches 289; Conservative 33; Mismatches 38; Indels 1; Gaps 1;

QY 1 MNKILFIFLTFSSVLTFPAVSADKIPGDSINIFGPRDRNESSPKHNLNHFAYSE 60

Db 1 MNKILFIFLTFSSVLTFPAVSADKIPGDSINIFGPRDRNESSPKHNLNHFAYSE 60

QY 61 SHTLYDRMTFLCLSSHTNLGACFTSENPSVSGETNITITQFTEKRLIKELQIKY 120
DB 61 SHLYDRMSFLCLSSQNTLNGACPSDDPATIDGETNITITQFTEKRLIKELQIKY 120
QY 121 KOLLFKSVKCPGLTUNSAHFNCNKAASGASLYLYIPAGELKNLPPGGIWDATLKLRYK 180
DB 121 KOFFLFRNANCPKSLALNSHFQCNQAGATLSLYIPAGELKNLPPGGVWNAVLKLNK 180
QY 181 RRYSEYTYGTYNITIKLTDKGNIQIWLPOFKSDARVDLNLRTGGGTIGRNSVDMCFY 240
DB 181 RRYDTTYGTYNITITNLDKGNIQIWLPOFKSNARVDLNLRTGGGTIGRNSVDMCFY 240
QY 241 DGYSTNSSSLEIRFQDNNPKSDGKPYLRKINDTKELIAYWLSLLACKSLTPPNTGTSINI 300
DB 241 DGYSTNSSSLEIRFQDNNPKSDGKPYLRKINDTKELIAYWLSLLACKSLTPNGOALNI 300
QY 301 ADAASLETNNRITAVTMPEISVPVLCWPGRLQLDKAVENPEAGQYMGNIIVTFPSSQT 360
DB 301 -NTASLETNNRITAVTMPEISVPVLCWPGRLQLDKAVENPEAGQYMGNIIVTFPSSSS 359
QY 361 L 361
DB 360 L 360

RESULT 2

US-08-483-101-16
; Sequence 16, Application US/08483101
; Patent No. 5932715
; GENERAL INFORMATION:
; APPLICANT: Scott, June R.
; APPLICANT: Froehlich, Barbara
; APPLICANT: Caron, Judy
; TITLE OF INVENTION: CS2 Proteins and Coding Sequences
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee and Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DCS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,101
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33878
; REFERENCE/DOCKET NUMBER: 6-95
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 363 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
US-08-483-101-16

Query Watch 52.4%; Score 988.5; DB 2; Length 363;
Best Local Similarity 55.0%; Pred No. 3, 5e-93;
Matches 204; Conservative 53; Mismatches 93; Indels 21; Gaps 11;

QY 3 KILFIPTLFFSVLTPFAVSADKIPGDE--STINIF-GPR-DNRESSPKHNILNNHITAY 58
DB 2 KIFIFL-----SIISAVVSRGRYBETTVGNLTGKFOAPRLDRSVQSPYINFTNHVAGY 57
QY 59 SRSHTLYDRMTFLCLSSHTNLGACFTSENPSVSGETNITITQFTEKRLIKELQIKY 117
DB 59 SLSHSLYDRIVFLCTSSSNPNVAGACPTI--GTSGVQYGTITITLQFTEKRLIKELQIKY 114
QY 118 KQYKOLLFKSVKCPG--LTINSAHFNCKNA-ASGASLYLYIPAGELKNLPPGGIWDAT 174
DB 115 AGNKKPIWENQCFDSNLMVLNSKSWSCGAHGNANGTLLNLYIPAGEINKLPFGGIWEAT 174
QY 175 LKLVRKRRYSET----YGYTITNITIKLTDKGNIQIWLPOFKSDARVDLNLRTGGGTI 230
DB 175 LILRLS-RGEVSVSTHYGNYTIVNITVLLDKGNIQWLPOFHSNPRVDNLRLPGNYKYS 233
QY 231 GRNSVDMCFYDGYSTNSSSLEIRFQDNNPKSDGKPYLRKINDTKELIAYWLSLLACKSL 290
DB 234 GNSLDMCFYDGYSTNSSSLEIRFQDNNPKSDGKPYLRKINDTKELIAYWLSLLACKSL 292
QY 291 TPNGTSLNIAADASLETNNRITAVTMPEISVPVLCWPGRLQLDKAVENPEAGQYMGNI 350
DB 293 YPVNGQSFTINDSSVLETNWRTAVMPEVNVPLCWPARLLLNADVNNPEAGQYMGNI 352
QY 351 NVTFTSSQTL 361
DB 353 KITFTSSQTL 363

RESULT 3

US-08-483-101-5
; Sequence 5, Application US/08483101
; Patent No. 5932715
; GENERAL INFORMATION:
; APPLICANT: Scott, June R.
; APPLICANT: Froehlich, Barbara
; APPLICANT: Caron, Judy
; TITLE OF INVENTION: CS2 Proteins and Coding Sequences
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee and Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,101
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33878
; REFERENCE/DOCKET NUMBER: 6-95
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 364 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-483-101-5

Query Match 49.2%; Score 927.5; DB 2; Length 364;
Best Local Similarity 48.1%; Pred. No. 6, 7e-87;
Matches 176; Conservative 65; Mismatches 118; Indels 7; Gaps 3;

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-617-697-10

Query Match 6.2%; Score 116.5; DB 2; Length 1600;
Best Local Similarity 21.6%; Pred. No. 0.024;
Matches 74; Conservative 60; Mismatches 130; Indels 79; Gaps 19;

QY 1 MNKILFTLFFSSVLFTFAVSADKIPGDSEIWNIFGRDRN--ESSPKENILNNEITAY 58
Db 1 LKKVIVLWFLSCQVYQGSWHYNVEAGSINKTESIGPIDR3AAASZPAHVI7HEHVAGY 60
QY 59 SE8HTLYDRMTFFCLSHNHTLNGACPTSENPPSSSVSGSTNITLOFTEKSLIKRELOIK 118
Db 61 NK0KSLFDRMTFLCMSSTDASKGACPTGSKSS--QGETNIKLI7FEKSLARKTLNLX 118
QY 119 GYKOLLPKSYNC---PSGLTNSAHFNCNKAASGASLYLVPAGELKNLPFGGIWDATL 175
Db 119 GYKFLYSDRCIHVYKXNLSNHTVCKVSGFRGVDFTLVYPOGEIDGLTGGINZATL 178
QY 176 KLRVKRYSEYTYTITIKLTDKGNIQIWLQPKFSADARVDLNLNRP7TGGTYIGNSV 235
Db 179 ELVRKHYDYNHGYKYNITVDLTDKGNIQVWTKFHSDFRIDLNL7PEGNKYGSGNVL 238
QY 236 DMCYDGYSTNSSLLEIRFQDNPKSDGKFLRKINDDTKEIATL7SLLAGKSLTNG 295
Db 239 EMCLYDGYSTHSQSIEMRFQDSQTNNEYNL7IKTGEPLKPLK7SLLGGREFYNN 298
QY 296 TSLNIAAASLETNWNRITATVMPSEISVPLVCPGRQLQDAKVENPAGQYMGNI7VFT 355
Db 299 EAFINTDSSFLFNWRLKSVLSPLQISPLVCPANLTFMSELN7PRAEYSGILN7VFT 358
QY 356 PSSOTL 361
Db 359 PSSSSL 364

RESULT 4
US-08-617-697-10
; Sequence 10, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: Of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,932
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1600 amino acids
; TYPE: amino acid
```

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-617-697-10

Query Match 6.2%; Score 116.5; DB 2; Length 1600;
Best Local Similarity 21.6%; Pred. No. 0.024;
Matches 74; Conservative 60; Mismatches 130; Indels 79; Gaps 19;

QY 48 HN-LNNHITAYSSHTLYDRMTFLC7SSNTLNGACPTSENPPSS-----SVSGETNITL 102
Db 816 YNEYSKH--AINSSHNL-----TILGNVTLCG-----ENSSSITCNTNITNKANVTL 862
QY 103 QF-----TEBSLJK3ELQIKGYKQLLFKSVNCP5GLTL-----NSAHFNON--- 144
Db 863 QATSNSTNGLKRTTLFLGNISVEGNLSLTCANANIVGNLSIAEDSTFKGEASDNLAITG 922
QY 145 ---KNAASGASLYLIPAGELKNLPFGGIWDATLKLVRKRYSEYTYGT-YTINTITKLT 200
Db 923 TFTNNGTANINIKGVKVLGDINK--CG-----LNTTNASGTQKTIINGNTI---NE 970
QY 201 KGNITQIWLQPKSDARVDL--NL7PTGGTYIGNSVDMCFYDGYSTNSSLLEIRFQ--- 255
Db 971 KGDLLN--KNIKADAE7QIGNISQKEGKNTL7ISSDKVNI-----TNQITIKAGVEGR 1022
QY 256 -DNNPKSDGKFLRKINDDTKEIATV7LSLLAG---XSL7PTNGTSLNIA7AASLETNW 311
Db 1022 SDSSEAFENANLTIQ-----TKELKLAGDLNLSGNKARITAKNGSDLTIGNASGNADAK 1076
QY 312 RITATVMPSEISVPLVCPFG-RLQLDKAVENPAGQYMGNI7VFT 353
Db 1077 K---V7FDKVKDSKITDGHNVTLN7SEVKTNGSSNAGNDNST 1116

RESULT 5
US-08-362-525-2
; Sequence 2, Application US/08362525
; Patent No. 6027910
; GENERAL INFORMATION:
; APPLICANT: KLS, FRANCISCUS M.
; APPLICANT: SCHREUDER, MAARTEN P.
; APPLICANT: TOSCHKA, HOLSER Y.
; APPLICANT: VERRIPS, CORNELIS T.
; TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ENZYMES TO THE
; TITLE OF INVENTION: CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DABBY & CUSHMAN, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: U.S.A.
; ZIP: 20005-3919
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,525
; FILING DATE: 04-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92202090.5
; FILING DATE: 08-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92203899.7
; FILING DATE: 14-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP93/01763
; FILING DATE: 07-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
```


103 QF-----TEKSLIKRE-QIKGYKOLLFKSVNCPGSLTNSAHFNCKNAASGASLY 154
791 QADTSNSNTGLKXRTLTGLNISVEGNLSUTGANAKVGNLSI--AEDSTFKGEAS----- 843
155 LYIPAGELKNI-PFGGIW--DATLKRVKRYSETVCTYT-----INITIKLT----- 199
844 -----DNLNITFTFNNGTANINIKGVVVKLQGDINNKGGLNITTNASTQXTIING 895
200 -----DKGNIQIWLPOFKSDARVDL--NLRPTGGGYIGRNSVDMGCFYDGYSTNSSLSLEIR 253
896 NITNEKGLNI--KNIKADAEIQIGNTISQKGNLTISDDKNI-----TNQITIKAG 946
254 FQ-----DNNPKSGDKFYRKINDDTKEIAYTUSLLAG---KSLTPNGTSLNIADAASL 306
947 VEGGRSDSEAEANLTIQ-----KELKLAGDLNISGFNKAEITAKGSDLTIGNASGG 1002
307 ETNWRITATVMPESVPLVCPG-RLQLDKVENPEAQYMGNIIVT 353
1002 NADAKK---VTFDKVKDSKISTDGHNVTLNSEVKTSGSSNAGNDNST 1046

Query Match 5.7%; Score 108; DB 4; Length 524;
Best Local Similarity 24.7%; Pred. No. 0.03; Mismatches 96; Indels 72; Gaps 18;
Matches 69; Conservative 42; Mismatches 96; Indels 72; Gaps 18;

US-09-242-913B-15
; Sequence 45, Application US/09242913B
; Patent No. 6551811
; GENERAL INFORMATION:
; APPLICANT: FONTAINE, THIERRY
; APPLICANT: FONTAINE, ROBERT
; APPLICANT: HARTLAND, ROBERT
; APPLICANT: MOUTNA, ISABELLE
; APPLICANT: LATGE, JEAN-PAUL
; TITLE OF INVENTION: METHOD FOR SORTING ANTIFUNGAL MOLECULES ACTING ON THE
; FILE REFERENCE: 05986-0007
; CURRENT APPLICATION NUMBER: US/09/242,913B
; CURRENT FILING DATE: 1999-10-13
; PRIOR APPLICATION NUMBER: PCT/FR97/01540
; PRIOR FILING DATE: 1997-08-29
; PRIOR APPLICATION NUMBER: 60/024,910
; PRIOR FILING DATE: 1996-08-30
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 15
; LENGTH: 524
; TYPE: PRF
; ORGANISM: Saccharomyces cerevisiae

US-09-242-913B-15

119 GYKOL--LFKSVNCPGSLTNSAHFNCKNAA---SGASLYLYIPAGELKNLPFGGIWDA 173
264 GYDKLNSTEDAVIP---LIFSEYGNKNTPTFEVSEGLY---GGLKNVPSGL--- 313
174 TLKLRVRYSETYGYTINITIKLTDKNIQIWLPOFKSD-ARVDLNR-----PTGGGT 228
314 -----VYEYTBANNYGL---VKLDDSGSL-----TYKDCFVNLESQKLNWSLPTTKBS 359
229 YIGRNSVDMC-----FYDGYTNSLSLE-----IRFQDNNPKSDGK----- 264
360 EISDSIYKCDNSAITNIYSGFGFNFLPSQPAETANWLEYGVNGINT-KGILTYAVP 418
265 ----FYLRKINDDTKEIAYTUSLLLAGSLTPTNGTSLNIADAASLFTNNR-ITAVTMP 319
419 TTNYYTIANNKDDT--ISATISYDKA-NSNELDVTATTYAKASNTSQSSRSLSSTSP 475
320 EISVPLVCPGRLQLDKVENPEAQYMGNI-NVTFTPS 357
476 SSSNGSSSTG-----SSASSSSKSKGVGNVNVPSFQS 510

Query Match 5.7%; Score 108; DB 4; Length 524;
Best Local Similarity 24.7%; Pred. No. 0.03; Mismatches 96; Indels 72; Gaps 18;
Matches 69; Conservative 42; Mismatches 96; Indels 72; Gaps 18;

US-09-242-913B-15

RESULT 11

US-09-206-942-45
; Sequence 45, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; FILE REFERENCE: 1038-861 MIS:jb
; CURRENT APPLICATION NUMBER: US/09/206,942
; CURRENT FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 09/167,568
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 45
; LENGTH: 1095
; TYPE: PRF
; ORGANISM: Haemophilus influenzae

Query Match 5.7%; Score 107; DB 4; Length 1095;
Best Local Similarity 24.6%; Pred. No. 0.12; Mismatches 71; Conservative 44; Mismatches 110; Indels 64; Gaps 15;

45 SPKXNLANHHTAYSESHLYD-----RMTFLCLSSHNTLNGACPTSENPSSSS 93
281 SIKPPIVEN--VHDGNETLFNGVSVLGGDVNFHFNASSNHWTHGVVIXSQNFNASE 337
94 VSGETNITLOFTKESLIKRE-QIKGYKOLLFKSVNCPGSLTNSAHFNCKNAAASGASL 153
333 GS-----SLRF-----KSGSTRAP--TIESDLTINATGGINSLNQVAGIDG 378
154 YLYIPAGELKNLPFGGIWDAITLKRVKRYSETYGYTI---NITIKLTDKNIQIWL 209
373 NLOKSIVANKNITFEG--GNITLAADKKPIBKGNITVKGANVTLRSANYGNDKSL- 434
210 QPKSDARVDLNRPTGGGYTIGRN-----SVDMCYDGYSTNSSLSLEIRFQDNNPKSD-- 262
435 SIRGNVTKGNLTVTGSAINIEKNLTVESGAKFLANPNYSFNVSGL-----FDNQGSNIS 490
263 ---GKFLRKINDDTKEIAYTUS-----LLLAGKSLTPTNGTSLNIAD 302
491 IAKGGAHFADIN-NYKSLNITNDSAYRTIIEG-NITNSNG-DANITD 536

US-09-206-942-45

RESULT 12
US-09-206-942-43
; Sequence 43, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; FILE REFERENCE: 1038-861 MIS:jb
; CURRENT APPLICATION NUMBER: US/09/206,942
; CURRENT FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 09/167,568
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 43
; LENGTH: 1101
; TYPE: PRF
; ORGANISM: Haemophilus influenzae

Query Match 5.7%; Score 107; DB 4; Length 1101;
Best Local Similarity 24.6%; Pred. No. 0.13; Mismatches 71; Conservative 44; Mismatches 110; Indels 64; Gaps 15;

US-09-206-942-43

QY 45 SPKGNILNHHITAYSESHTLYD-----RMTFLCLSHNTLNGACPTSPNPSSSS 93
DB 287 SIKPPIVSN--VHDGNHTLNGRVSVLGGDVNHFENASSNHWTHGWIKSQNFNASE 343
QY 94 VSGETNITLQTEKSLIKRELOIKGYKQLLPKSVNCPGLTASAHFNCKNAASGASL 153
DB 344 GS-----SLRF-----KSEGSTRTAF--TIESDTLNATGNSISUNQVAGIDG 384
QY 154 YLYTPAGELKXLPFGGIWDATLKRVRKRYSETGYTI---NITIKLTGKNIQIWL 209
DB 395 NLQKSLVANKNITPFG---GNITLADKKPIE-KGNIIVKEGANVTLASVGNCKSAL- 440
QY 210 QPKSDARYDLNLKRPFGGTYIGRN-----SVDVCFYDGYSTNSSSLEIRFQDNPKSD- 262
DB 441 SIRGNVTNKGNTLVGTGSAINTIEKLTVEGSAKFLANPNYSFNVSGI-----FDNQGKSNIS 496
QY 263 ---GKPYLRKINDDTKEIAYTIS-----LLLAGKSLPTNGTSLNIAD 302
DB 497 IAKGGAHFKDIN-NTKSLNITNTSDAYRTIEG-NITNSNG-DLNIYD 542

RESULT 13

US-08-737-716-13

; Sequence 13, Application US/08737716

; Patent No. 5955258

; GENERAL INFORMATION:

; APPLICANT: Gierbe BUIST

; APPLICANT: Gerard VENEMA

; APPLICANT: Jan KOK

; APPLICANT: Adrianus Marinus LEDEBOER

; TITLE OF INVENTION: Process for the lysis of a culture of lactic

; TITLE OF INVENTION: acid bacteria by means of a lysin, and uses of the resulting

; TITLE OF INVENTION: lysed culture.

; NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:

; ADDRESSES: Pillsbury Madison & Sutro, L.L.P.

; STREET: 1100 New York Avenue, N.W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20005-3918

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Microsoft Word

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/737,716

; FILING DATE: 22-APR-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/NL95/00170

; FILING DATE: 12-MAY-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: EP 94201353.3

; FILING DATE: 12-MAY-1994

; INFORMATION FOR SEQ ID NO: 13:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 671 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; ORIGINAL SOURCE:

; ORGANISM: Streptococcus faecalis

; IMMEDIATE SOURCE:

; CLONE: Fig.5a (S. faecalis)

US-08-737-716-13

Query Match

Best Local Similarity 5.6%; Score 106; DB 2; Length 671;

Matches 73; Conservative 39; Mismatches 103; Indels 94; Gaps 17;

QY 41 RVNESPKNIL-LANNHITAYSESHTLYDHTFLCLSSHNTLNGACPTSPNPSSSV- 94
DB 309 RYATDFSNKLNHNVITAY--NLQYDTPSSGGNIGGTVNAPGTGGNNQSGTNYIYTVK 366
QY 95 SCET--NITLQF-----TEKRL--IKRELQ-KGYKQLLPKSVNCPGSLTASAHFNCKN 146
DB 367 SGDILNKIAAQQYGVSVANLRSMNGISGDLIFVGOKLIVKGA--SGVTGGSGNGSKNN 423
QY 147 AASGASLYLYTPAGELKXLPFGGIWDATLKRVRKRYSETGYTI---NITIKLTGKNIQI 206
DB 424 -QSGTNTVYTVKSGDTLN-----KIAAQY-----VTV-----ANLRS 455
QY 207 WLPQKSDARYDLNL-----RPTGGGTYIGRNSVDMCFYDGYSTNSSSLEIRFQDN 258
DB 456 W-----NGISGDLIFVGOKLIVKGTSGNT-----GSSNGGS-----NNN 492
QY 259 PKSDGKPYLRKINDDTKEIA--YTIS-----LLLAGKSLPTNGTSLNIADA 303
DB 492 QSGTNTVYTVKSGDTLNKIAAQQYGVSVANLRSMNGISGDLIFAGOKIIVKGTSGNTG 551
QY 304 ASLETNWN 312
DB 553 SNGGSMNQ 560

RESULT 14

US-09-206-942-35

; Sequence 35, Application US/09206942

; Patent No. 6432669

; GENERAL INFORMATION:

; APPLICANT: Loosmore, Sheena M.

; APPLICANT: Yang, Yan-Ping

; APPLICANT: Klein, Michel H.

; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High

; TITLE OF INVENTION: Molecular Weight Proteins

; FILE REFERENCE: 1038-861 MIS.1b

; CURRENT APPLICATION NUMBER: US/09/206,942

; CURRENT FILING DATE: 1998-12-08

; EARLIER APPLICATION NUMBER: 09/167,568

; EARLIER FILING DATE: 1998-10-07

; NUMBER OF SEQ ID NOS: 95

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 35

; LENGTH: 915

; TYPE: PRT

; ORGANISM: Haemophilus influenzae

US-09-206-942-35

Query Match 5.4%; Score 101; DB 4; Length 915;

Best Local Similarity 21.1%; Pred. No. 0.38;

Matches 94; Conservative 48; Mismatches 129; Indels 138; Gaps 19;

QY 21 VSADKIPGDESITNIFGPR---DRENSPKH-----NIIIN-----NHITAYS 59
DB 252 ISSDKVNTERTITIKAGVNGSDNSDNEATSNLTIKTELKLTNDLNSGPNKARITAKD 311
QY 60 ESH-TLYD-----RMTF-----LCUSSN-TLNGACPTSPNPSSSVSGETNI 100
DB 312 NGNLTIGDSDAGNTDAKKVTPSNVKDGSISASDENVTINSKVETSGDSTDGCGNNNT 371
QY 101 TLQFTKESLIRELOIKGYKQL---LFKSVNCPGSLTASAHFNCKNAASGASLYLYI 157
DB 372 GLTITAKNVTVN--NITSHKTVNTASENVTVKAGTIN-----AATGVEVTA 419
QY 153 PAGELKXLPFGGIWDATLKRVRKRYSETGYTI---NITIKLTGKNIQIWLFPQKSDARV 217
DB 420 KTGDIK---GGI-----EBSNGVNI----- 437
QY 219 DLNLRPTGGTYI-----GRNSVDMCFYDGYSTNSSSLEIRFQDNPKSDGKPYLRKIN 271
DB 438 -----TASGDTLNVSNITGQN-VTAAASGAVTTKGTSTNATTGNANITTK--TGEIN 488
QY 272 DDTKFIATLSLLLAGKSLTFTNGISLNIAADASLETNWNRTAVTMPPEISFVILCWPCR 331

Db 489 GEVKSAGNVNITASGNTLVNSNITQNVTVTAN-----SGAITTEGST----- 533
QY 332 LQDAKVENPEAGQYMGNN-----VTFTPSQTL 361
Db 534 --INATTGDANITQTGNGINGKVESSGSVTLIATGQTL 570

RESULT 15
US-09-206-942-37
; Sequence 37, Application US/09206942
; Patent No. 6432569
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; TITLE OF INVENTION: Molecular Weight Proteins
; FILE REFERENCE: 1038-861 MIS:Jb
; CURRENT APPLICATION NUMBER: US/09/206,942
; CURRENT FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 09/167,568
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 1222
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-206-942-37

Query Match 5.4%; Score 101; DB 4; Length 1222;
Best Local Similarity 21.1%; Pred. No. 0.61;
Matches 84; Conservative 48; Mismatches 129; Indels 138; Gaps 19;
QY 21 VSADKIPGESITNIFGPR----DRNESSPKH-----NILN-----NHITAYS 59
Db 559 ISSDKVNITERITIKAGVNGDSDNEATSANLTIKELKLTNDLNI SGFNKAEITAKD 618
QY 60 ESH-TLYD-----RMTR-----LCLSSHN-TLNGACPTSENPSSSSVSGETNI 100
Db 619 NSKLTIGDSDAGNTDAKKVTFNSVNDKSKISASHNVTLNSKVETSGDTEDEGGNNNT 678
QY 101 TLQFTKRSLIKRELIQKGYQL---LFKSVNGCPGSLTNSAHFNCKNAAGSASYLYI 157
Db 679 GLTIITAKNVTVNN--NITSKTVNITASENVTTKAGTTN-----ATTGSVEVTA 726
QY 158 PAGELXNLPFGGIWDATLKURVKRISFTGYVTINIKITDKGNLIQWLPOFKSDARV 217
Db 727 KTGDIK-----GGI-----ESNSGNVNI----- 744
QY 218 DLNLRPTGGTYI-----GENSVDMCFYDGYSTNSSSLZIRFQDNNPKSDGKPYLRKIN 271
Db 745 -----TASGDTLVNSNITQCN-VTVAASAGAVTTKGTSTINATTGNANITTK--TGEIN 795
QY 272 DDTKEIAYTLLSLAGKSLTPTNGTSLNIADAASLETNWNRIATVMPPEISVPLCWPCR 331
Db 796 GEVKSAGNVNITASGNTLVNSNITQNVTVTAN-----SGAITTEGST----- 840
QY 332 LQDAKVENPEAGQYMGNN-----VTFTPSQTL 361
Db 841 --INATTGDANITQTGNGINGKVESSGSVTLIATGQTL 877

Search completed: May 6, 2004, 10:20:19
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

QM protein - protein search, using sw model

Run on: May 6, 2004, 10:19:17 ; Search time 48 Seconds
(without alignments)
2087.533 Million cell updates/sec

Title: US-09-839-894-10
Perfect score: 1886
Sequence: 1 MNKILFIFLFFSSVLFFA.....EAGQVMGNINVTFFSSQTL 361

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277565755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/1/pubpaa/CT_NEW_PUB.pep.*
3: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/prodata/1/pubpaa/CTUS_PUBCOMB.pep.*
7: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/prodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/prodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/prodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pep.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1886	100.0	361	9	US-09-839-894-10
2	1862.5	98.8	361	9	US-09-839-894-28
3	1542	81.8	359	9	US-09-839-894-29
4	937	49.7	362	9	US-09-839-894-31
5	916.5	48.6	364	9	US-09-839-894-30
6	247.5	13.1	353	9	US-09-839-894-32
7	120	6.4	1335	12	US-10-282-122A-42853
8	119.5	6.3	1770	9	US-09-303-232-2
9	116.5	6.2	1600	13	US-10-092-880-10
10	114	6.0	650	15	US-10-369-493-22177
11	113.5	6.0	513	12	US-10-282-122A-61451
12	111.5	5.9	1005	14	US-10-193-764-41
13	111.5	5.9	1021	14	US-10-193-764-39
14	109	5.8	2383	9	US-09-912-020-302
15	109	5.8	2383	12	US-10-282-122A-42852

16	108	5.7	524	14	US-10-347-278-15	Sequence 15, Appl
17	108	5.7	524	14	US-10-347-252-15	Sequence 15, Appl
18	108	5.7	2529	12	US-10-282-122A-58834	Sequence 58834, A
19	105	5.6	598	12	US-10-282-122A-60611	Sequence 60611, A
20	103.5	5.5	691	15	US-10-369-493-23473	Sequence 23473, A
21	103.5	5.5	2399	12	US-10-282-122A-59054	Sequence 59054, A
22	101	5.4	915	14	US-10-193-764-35	Sequence 35, Appl
23	101	5.4	1222	14	US-10-193-764-37	Sequence 37, Appl
24	101	5.4	1228	14	US-10-193-764-34	Sequence 34, Appl
25	101	5.4	1599	13	US-10-092-880-9	Sequence 9, Appl
26	100.5	5.3	1004	14	US-10-193-764-53	Sequence 53, Appl
27	100.5	5.3	1010	14	US-10-193-764-51	Sequence 51, Appl
28	99.5	5.3	969	14	US-10-193-764-32	Sequence 32, Appl
29	99.5	5.3	975	14	US-10-193-764-30	Sequence 30, Appl
30	99.5	5.3	1300	12	US-10-282-122A-67412	Sequence 67412, A
31	99	5.2	839	10	US-09-738-269-23	Sequence 23, Appl
32	99	5.2	839	13	US-10-023-437-23	Sequence 23, Appl
33	99	5.2	992	14	US-10-193-764-57	Sequence 57, Appl
34	99	5.2	998	14	US-10-193-764-55	Sequence 55, Appl
35	99	5.2	1848	10	US-09-839-996-6	Sequence 6, Appl
36	99	5.2	1848	12	US-10-645-655-6	Sequence 6, Appl
37	99	5.2	1848	14	US-10-080-505-6	Sequence 43, Appl
38	97	5.1	977	14	US-10-193-764-49	Sequence 49, Appl
39	97	5.1	983	14	US-10-193-764-47	Sequence 47, Appl
40	96.5	5.1	627	12	US-10-425-114-62687	Sequence 62687, A
41	95.5	5.1	925	14	US-10-101-464A-922	Sequence 922, App
42	95.5	5.1	1194	12	US-10-282-122A-46577	Sequence 46577, A
43	95	5.0	867	9	US-09-839-894-6	Sequence 6, Appl
44	94.5	5.0	312	12	US-10-424-599-234245	Sequence 234245
45	94.5	5.0	2353	9	US-09-797-862-33	Sequence 33, Appl

ALIGNMENTS

RESULT 1
US-09-839-894-10
; Sequence 10, Application US/09839894
; Patent No. US20040176868A1
; GENERAL INFORMATION:
; APPLICANT: Altboum, Zeev
; APPLICANT: Barry, Eileen M.
; APPLICANT: Levine, Myron M.
; APPLICANT: University of Maryland
; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
; TITLE OF INVENTION: CSA OPERON
; FILE REFERENCE: UOFMD.006A
; CURRENT APPLICATION NUMBER: US/09/839,894
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/198,626
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 361
; TYPE: PRT
; ORGANISM: E. coli
US-09-839-894-10

Query Match	100.0%;	Score	1886;	DB	9;	Length	361;
Best Local Similarity	100.0%;	Pred. No.	3.3e-181;				
Matches	361;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	MNKLIFLFFSSVLFFAVSADKIPGDSINIFGPRDNSSSPKHNINNHITAYSE	60				
Db	1	MNKLIFLFFSSVLFFAVSADKIPGDSINIFGPRDNSSSPKHNINNHITAYSE	60				
Qy	61	SHLYDRMTFLCSSHNTLNAGPTSENPPSSSVSGTNTITLOFTKRSIRKRELQIKGY	120				
Db	61	SHLYDRMTFLCSSHNTLNAGPTSENPPSSSVSGTNTITLOFTKRSIRKRELQIKGY	120				
Qy	121	KQLIFSVNCPGSLTNSAHFNCKNAASGASLYIYPAGELKNLPFGGTWDATLKRVK	180				

```
Db 121 KQLLPKSVNCPSSLTNSAHFNCNKAASGLYLYIPAGELKNLPFGGIWDATLKLVRK 180
QY 181 RYSETYGTYYTINITIKLTDKGNIOIWLPOFKSDARVDNLNRPCTGGTYIGRNSVDMCFY 240
Db 181 RYSETYGTYYTINITIKLTDKGNIOIWLPOFKSDARVDNLNRPCTGGTYIGRNSVDMCFY 240
QY 241 DGYSTNSSSLEIRFQDNPKSDGKFLRKINDDTKEIAYTSLLSLLAGKSLTPNGTSLNI 300
Db 241 DGYSTNSSSLEIRFQDNPKSDGKFLRKINDDTKEIAYTSLLSLLAGKSLTPNGTSLNI 300
QY 301 ADAASLETNNWRITAVTMPPIISVPVLCWPGRLQIDAKVENPEAGQYMGNIIVTFPSSQT 360
Db 301 ADAASLETNNWRITAVTMPPIISVPVLCWPGRLQIDAKVENPEAGQYMGNIIVTFPSSQT 360
QY 361 L 361
Db 361 L 361

RESULT 2
US-09-839-894-28
; Sequence 28, Application US/09839894
; Patent No. US20020176868A1
; GENERAL INFORMATION:
; APPLICANT: Altboum, Zeev
; APPLICANT: Barry, Eileen M.
; APPLICANT: Levine, Myron M.
; APPLICANT: University of Maryland
; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
; FILE REFERENCE: UOFMD.006A
; CURRENT APPLICATION NUMBER: US/09/839,894
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 2001-04-20
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ETEC Protein Homology Sequence
US-09-839-894-28

Query Match 98.8%; Score 1862.5; DB 9; Length 361;
Best Local Similarity 99.4%; Pred. No. 7.8e-179;
Matches 359; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 MNKILFIPTLFFSSVLTFAVSADKIPGDESITNIFGPRDRNESSPKHNILNNHITAYSE 60
Db 2 MNKILFIPTLFFSSVLTFAVSADKIPGDESITNIFGPRDRNESSPKHNILNNHITAYSE 61
QY 61 SHLYDRMTFCLSHNTLNACPTSENPPSSSVSGETNITLQFTEKSLIKRELQIKGY 120
Db 62 SHLYDRMTFCLSHNTLNACPTSENPPSSSVSGETNITLQFTEKSLIKRELQIKGY 121
QY 121 KQLLPKSVNCPSSLTNSAHFNCNKAASGLYLYIPAGELKNLPFGGIWDATLKLVRK 180
Db 122 KQLLPKSVNCPSSLTNSAHFNCNKAASGLYLYIPAGELKNLPFGGIWDATLKLVRK 181
QY 181 RYSETYGTYYTINITIKLTDKGNIOIWLPOFKSDARVDNLNRPCTGGTYIGRNSVDMCFY 240
Db 182 RYSETYGTYYTINITIKLTDKGNIOIWLPOFKSDARVDNLNRPCTGGTYIGRNSVDMCFY 241
QY 241 DGYSTNSSSLEIRFQDNPKSDGKFLRKINDDTKEIAYTSLLSLLAGKSLTPNGTSLNI 300
Db 242 DGYSTNSSSLEIRFQDNPKSDGKFLRKINDDTKEIAYTSLLSLLAGKSLTPNGTSLNI 300
QY 301 ADAASLETNNWRITAVTMPPIISVPVLCWPGRLQIDAKVENPEAGQYMGNIIVTFPSSQT 360
Db 301 ADAASLETNNWRITAVTMPPIISVPVLCWPGRLQIDAKVENPEAGQYMGNIIVTFPSSQT 360
```

```
QY 361 L 361
Db 361 L 361

RESULT 3
US-09-839-894-29
; Sequence 29, Application US/09839894
; Patent No. US20020176868A1
; GENERAL INFORMATION:
; APPLICANT: Altboum, Zeev
; APPLICANT: Barry, Eileen M.
; APPLICANT: Levine, Myron M.
; APPLICANT: University of Maryland
; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
; FILE REFERENCE: UOFMD.006A
; CURRENT APPLICATION NUMBER: US/09/839,894
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/198,626
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ETEC Protein Homology Sequence
US-09-839-894-29

Query Match 81.8%; Score 1542; DB 9; Length 359;
Best Local Similarity 80.9%; Pred. No. 1.6e-146;
Matches 292; Conservative 32; Mismatches 35; Indels 2; Gaps 2;

QY 1 MNKILFIPTLFFSSVLTFAVSADKIPGDESITNIFGPRDRNESSPKHNILNNHITAYSE 60
Db 1 MNKILFIPTLFFSSVLTFAVSADKIPGDESITNIFGPRDRNESSPKHNILNNHITAYSE 60
QY 61 SHLYDRMTFCLSHNTLNACPTSENPPSSSVSGETNITLQFTEKSLIKRELQIKGY 120
Db 61 SHLYDRMTFCLSHNTLNACPTSENPPSSSVSGETNITLQFTEKSLIKRELQIKGY 120
QY 121 KQLLPKSVNCPSSLTNSAHFNCNKAASGLYLYIPAGELKNLPFGGIWDATLKLVRK 180
Db 121 KQFLFKNANCPSKIALNSSHQCNREQASGATISLYIPAGELKNLPFGGVNNAVLKLVK 180
QY 181 RYSETYGTYYTINITIKLTDKGNIOIWLPOFKSDARVDNLNRPCTGGTYIGRNSVDMCFY 240
Db 181 RYR-TYGTYYTINITIKLTDKGNIOIWLPOFKSNARVDNLNRPCTGGTYIGRNSVDMCFY 239
QY 241 DGYSTNSSSLEIRFQDNPKSDGKFLRKINDDTKEIAYTSLLSLLAGKSLTPNGTSLNI 300
Db 240 DGYSTNSSSLEIRFQDNPKSDGKFLRKINDDTKEIAYTSLLSLLAGKSLTPNGTSLNI 299
QY 301 ADAASLETNNWRITAVTMPPIISVPVLCWPGRLQIDAKVENPEAGQYMGNIIVTFPSSQT 360
Db 300 -NTASLETNNWRITAVTMPPIISVPVLCWPGRLQIDAKVENPEAGQYMGNIIVTFPSSQT 358
QY 351 L 361
Db 359 L 359

RESULT 4
US-09-839-894-31
; Sequence 31, Application US/09839894
; Patent No. US20020176868A1
; GENERAL INFORMATION:
; APPLICANT: Altboum, Zeev
; APPLICANT: Barry, Eileen M.
; APPLICANT: Levine, Myron M.
; APPLICANT: University of Maryland
```

1 TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
2 TITLE OF INVENTION: CSA OPERON
3 FILE REFERENCE: UOFMD.006A
4 CURRENT APPLICATION NUMBER: US/09/839,894
5 CURRENT FILING DATE: 2001-04-20
6 PRIOR APPLICATION NUMBER: 60/198,626
7 PRIOR FILING DATE: 2000-04-20
8 NUMBER OF SEQ ID NOS: 40
9 SOFTWARE: FastSeq for Windows Version 4.0
10 SEQ ID NO 31
11 LENGTH: 362
12 TYPE: PR1
13 ORGANISM: Artificial Sequence
14 FEATURE:
15 OTHER INFORMATION: ETEC Protein Homology Sequence
16 US-09-839-894-31

Query Match 49.7%; Score 937; DB 9; Length 362;
Best Local Similarity 53.1%; Pred. No. 1.7e-85;
Matches 197; Conservative 54; Mismatches 98; Indels 22; Gaps 12;
1 QY 3 KILFIPTLPSSVLFPAVSADKIPGDE--SIITNIF-GER-DRNESSPKHNIINNHITAY 58
2 Db 2 KILFIPL----SIIFSADVSAAGYPTTGVGNLTKSQAQLRDSVQSPIYFTNIVAGY 57
3 QY 59 SESHLYDRMTFLCLSSHTNLGACPTSENPPSSSVS-GETNITLQFTKRSLIKRELQI 117
4 Db 58 SLASHSLYDAIVFCTSSNPVNGACPTI---GTSGVQYGTITLQFTKRSLIKENINI 114
5 QY 118 KGYKQLLFKSVNCPG--LTLSAHFNCKNA--ASGASLYLYIPAGELKNLPFGGWDAT 174
6 Db 115 AGNKKPIWENQSCDFNIMVLSKSCGANGTILNMLYIPAGEINKLPFGGWEAT 174
7 QY 175 LKLRVKRRYSET----YGIYTNITIKLTDKGNIQIWLPOFKSDARVDLNLRPFGGTYI 230
8 Db 175 LILRLS-RGEVSSYTHGYNTVNIITVDLTDKGNIQWLPGFHSNPRVDLNLRIQNVKYS 233
9 QY 231 GRNSVDMCFYDGYSTNSSLSLEIRFQDNNPKSDGKFLYLRKINDTKETIAYTLSSLACKSL 290
10 Db 234 GSNLDMCFYDGYSTNSDSVMVIFQDNPNTSSEYNLYKIG-GTEKLPYAVS-LIGBKIF 291
11 QY 291 TPNGTSLNLTADAASLETNWNRITAVTMPEISVPLVCWPGRLQIDAKVENPEAGQYMGNI 350
12 Db 292 YPVNGQSFINDSSVLETNWNRITAVAMPENNVPEVLCWPARLLNADVNAPDAGQVSGQI 351
13 QY 351 NVFTPTSSQTL 361
14 Db 352 YIIFTPSEVNL 362

RESULT 5
US-09-839-894-30
; Sequence 30, Application US/09839894
; Patent No. US20020176868A1
; GENERAL INFORMATION:
; APPLICANT: Altboum, Zeev
; APPLICANT: Barry, Eileen M.
; APPLICANT: Levine, Myron M.
; APPLICANT: University of Maryland
; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
; FILE REFERENCE: UOFMD.006A
; CURRENT APPLICATION NUMBER: US/09/839,894
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/198,626
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 364
; TYPE: PR1
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ETEC Protein Homology Sequence
; US-09-839-894-31

1 OTHER INFORMATION: ETEC Protein Homology Sequence
2 US-09-839-894-30
3 Query Match 48.6%; Score 916.5; DB 9; Length 364;
4 Best Local Similarity 48.1%; Pred. No. 2e-83;
5 Matches 176; Conservative 63; Mismatches 120; Indels 7; Gaps 3;
6 QY 1 MKKILFIPTLPSSVLFPAVSADKIPGDESIITNIFGPRDRN--BSSPKHNIINNHITAY 58
7 Db 1 MKKVIFVLSMFLCSQVYQSWHTNVEAGSINTFSGIDRDAASYPDAHYFHFHEVAGY 60
8 QY 59 SESHLYDRMTFLCLSSHTNLGACPTSENPPSSSVSEETITLQFTKRSLIKRELQI 118
9 Db 61 NKDHSLSFDRMTFLCMSSIDASKGACPTGNSKSS--QGETNIKLIPTKRSKLARKTILNK 118
10 QY 119 GYKQLLFKSVNCPG--PSGLTNSAHFNCKNAASGASLYLYIPAGELKNLPFGGIWDATL 175
11 Db 119 GYKRFLYESDRCIHYVDKMLNLSHTVKCVGSTRGVDFTLIYPOGEIDGLTGGIWKATL 178
12 QY 175 KLRVKRRYSETYCTYTNITIKLTDKGNIQIWLPOFKSDARVDLNLRPFGGTYIGNSV 235
13 Db 179 ELRVXRHYDYNHGTYKVNITVDLTDKGNIQVWTPKPHSDPRIDLNLRPNGKNGKYSNVL 238
14 QY 236 DMCYFDGYSTNSSLSLEIRFQDNNPKSDGKFLYLRKINDTKETIAYTLSSLACKSLPTMG 295
15 Db 239 EMCLYDGYSTHSQSIEMRQDDSDQTCNNEVNLKTGEP.LKLPYK-SILLGGRFPYNG 298
16 QY 296 TSLNTADAASLETNWNRITAVTMPEISVPLVCWPGRLQIDAKVENPEAGQYMGNI 355
17 Db 299 KAPTINDTSSLFINWNRKSVSLPQISIPVLCWPAANLTFMSELNPEAGEYSGLNVTF 358
18 QY 356 PSSQTL 361
19 Db 359 PSSSSL 364

RESULT 6
US-09-839-894-32
; Sequence 32, Application US/09839894
; Patent No. US20020176868A1
; GENERAL INFORMATION:
; APPLICANT: Altboum, Zeev
; APPLICANT: Barry, Eileen M.
; APPLICANT: Levine, Myron M.
; APPLICANT: University of Maryland
; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
; FILE REFERENCE: UOFMD.006A
; CURRENT APPLICATION NUMBER: US/09/839,894
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/198,626
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 353
; TYPE: PR1
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ETEC Protein Homology Sequence
; US-09-839-894-32
Query Match 13.1%; Score 247.5; DB 9; Length 353;
Best Local Similarity 28.9%; Pred. No. 5.5e-16;
Matches 89; Conservative 42; Mismatches 126; Indels 51; Gaps 13;
QY 77 NTANG-ACPTSENPPSSS-----VSGETNITLQFTKRSLIKRELQIKGYKQLLFK 126
Db 72 NVLGGWVCRNRNNEGCEETHLVWYAFGAYSILRFRQISHAEITL----- 120
QY 127 SVNCPGLTNSAHFNCPG--NKNAAAS---GASLYLYIPAGELKNLPFGGIWDATLKL-R 178
Db 121 -----ILLGSVRDACTGVINWNAACQGRSLKLRIPSEELAKIPTSGTWKATLVLDY 173

QY 179 VKRRYSETGYTINTIKLTD--KGNIOIWLPOF-KSDARVDLNLRLPTGGGTYIGNSV 235
Db 174 LQGGDDPPJGTSTDTLNLVDHFAKNAALYFQFGTATPRVDLNLHRNWSQMSGRANL 233
QY 236 DMCYDGYSTNSSLSIRFQDNPKSDGKFLYLRKINDTKETAYTSLLLAGKSLPTNG 295
Db 234 DMLCYDC-GYKARSLQW--MEGSKSGTGFGVIK-SDSADTIDYAVSMYGGRSIPVTRG 289
QY 296 TSLNIADAASLETNWNRITAVTMEISVPVLCWPGRLQLDLAK---VENPEAGGYMGNIW 352
Db 290 VEFSLNDVDKAATR-----PWLPGQRAVRCVPEVLTTLTTPQFNIREKRSGBEYQGLTV 344
QY 353 TETPSSQT 360
Db 345 TMLMGTOI 352

RESULT 7
US-10-282-122A-42853
; Sequence 42853, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EUTRA.034A
; CURRENT APPLICATION NUMBER: US/10/282-122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42853
; LENGTH: 1335
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-282-122A-42853

Query Match 6.4%; Score 120; DB 12; Length 1335;
Best Local Similarity 20.4%; Pred. No. 0.031;
Matches 80; Conservative 66; Mismatches 156; Indels 90; Gaps 15;
QY 13 SSVLFTFAVSADKIPGDESITNIFGPRDRNESSPKHNLNHTAY---SESHLYDRMT 69

Db 746 SAKIATLSASNGVLANENAAVTSVNVVADEGS---NPINDHTVTFVAVLSGSATSFNNQN 802
QY 70 FCLSSHNTLNGACPTSENPSSSVSGEINITLQTEKRSLLI-----K 112
Db 803 ----TAKTVDNGLA-TFDLKSSKQEDNTVEVILENGVKQIILVSVFVGDSSTAQVDLOKSK 857
QY 113 RELQIKGV-KOLLFKSVNCPSGLTLSAHSFNCNKAASGASILYIIPAGELKNLPFGGIW 171
Db 858 NEVVADGNDSATMTA-VRDACKGNLLNDVKVTNVNSAAAKLSQTEVNSHD-----GIA 910
QY 172 DATLKURVKRRISSETGYTINTIKLTDKGNIOIWLPOFKSDARVDLNLRLPTGGGTYIG 231
Db 911 TATLT-----SLKNGDYTVTASVSSGQANQOVIFIGDQSTAALTLSV-PSGDIIV-- 960
QY 232 RNSVDMCFYDGYSTNSSLE---IRFQDNPKSDGKFLYLRKINDTKETAYTSLLLAG 287
Db 961 -----TNTAPLHMTATLQDKNGNPLKDKETFSVEND-----VASRFSISNSG 1003
QY 288 KSLTPTNGTSL-----NIADAASLETNWNRITAVTMEISVPVLC 327
Db 1004 KGMTDSNGTAIASLTGTLAGTHMTARLANSNVSDTQPMTFVADKDRVAVVVLQTSKAEII 1063
QY 328 WPG--RLQLDKAVENPEAGGYMGNIWTFPS 357
Db 1064 GNCVDETTLTATVKDP-FDNVVKLSVWFRFS 1094

RESULT 8
US-09-303-232-2
; Sequence 2, Application US/09303232A
; Patent No. US20020006657A1
; GENERAL INFORMATION:
; APPLICANT: Bayer Aktiengesellschaft
; TITLE OF INVENTION: Nucleic Acids which encode
; TITLE OF INVENTION: Insect acetylcholine receptor subunits
; FILE REFERENCE: Je A 33 020-Foreign Countries
; CURRENT APPLICATION NUMBER: US/09/303,232A
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: DE 198 19 829.9
; EARLIER FILING DATE: 1998-05-04
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 770
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-303-232-2
Query Match 6.3%; Score 119.5; DB 9; Length 770;
Best Local Similarity 20.4%; Pred. No. 0.015;
Matches 83; Conservative 45; Mismatches 140; Indels 135; Gaps 20;
QY 9 TLFFSSVLFTFAVSADKIPGDESITNIFGPRDRNESSPKHNLNHTAY---SESHLYDRMT 54
Db 249 TIAFISYLGSAFALQKLNSSSSSSSSN-----SSNNSSTQILNGLNKGKSWIFLLIYLNLSAK 304
QY 55 --JTAVSESHLYDRMTFCLSSHNTLNGACPTSENPSSSVSGEINITLQTEKRSLLI 112
Db 305 VCLAGVHEKLLHD-----LLDPNITLERPVLESPLQLSPGLTLMQIIDVDKQNLV 359
QY 113 RELQIKGVKOLLFKSVNCPSGLTLSAHSFNCNKAASGASILYIIPAGELKNLPFGGIW 172
Db 360 TNVWLK-----LEWDMNLRWNTSDYGGVX-----DLR-IPPHRIWK 395
QY 173 ATLKLVRKRRYSET-----YGYTYTINTIKLTDKGNIOIWLPO--FKSDARVDLNLRLPTGG 226
Db 396 PDVLM-----YNSADEGFDGTQYNVVVR--NNGSC--LYVPPGIFKTKCIDITWFP--- 444
QY 237 GTYIGRNSVDMCF-----YDGYSTNSSLSIRFQDNPKSDGKFLYLRKINDTKETAY 276
Db 445 ---FDDQRCMKFGSTWYDGF-----QLDIQLQDETGDIDSSVYVINGEWELLGVPGKRNE 496

QY 277 IAYT-----LSLLACKSLTPNGTSLNIADA 303
 Db 497 IYVCCPEYIDITFAIIRRTLYFFNLIPCVLIASMLLGFLLPDSGEXLSLGT 556
 QY 304 ASLE-TWNR-TATMPIS-VFV-----LCW-----PCR 331
 Db 557 ILLSLTVFLNVAETMPATSDAVPLWIRIVFLCWLPWILNRSRQR 602

RESULT 9

US-10-092-880-10
 ; Sequence 10, Application US/10092880
 ; Publication No. US20020164354A1

GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J.
 ; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE
 ; TITLE OF INVENTION: HAEMOPHILUS
 ; FILE REFERENCE:
 ; CURRENT APPLICATION NUMBER: US/10/092,880
 ; CURRENT FILING DATE: 2002-03-08
 ; PRIOR APPLICATION NUMBER: 09/155,614
 ; PRIOR FILING DATE: 1998-09-30
 ; PRIOR APPLICATION NUMBER: 08/617,697
 ; PRIOR FILING DATE: 1996-04-01
 ; PRIOR APPLICATION NUMBER: PCT/US97/04707
 ; PRIOR FILING DATE: 1997-04-01
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: PatentIn ver. 2.1
 ; SEQ ID NO 10
 ; LENGTH: 1600
 ; TYPE: PRT
 ; ORGANISM: Haemophilus influenzae
 US-10-092-880-10

Query Match 6.2%; Score 116.5; DB 13; Length 1600;

Best Local Similarity 21.6%; Pred. No. 0.092; Mismatches 130; Indels 79; Gaps 19;
 Matches 74; Conservative 60;

QY 48 HNILNNHITAYGESHLYDRMTPLCLSSHNLTNGACPTSENPSS-----SVSGETNITL 102
 Db 816 YNEYSKH-AINSHNL-----TLGGNVTLGG-----ENSSSITGMINITNKANVTL 862
 QY 103 QP-----TEKRSILKELOIKGVKQLLFKSNVCPGLTL-----NSAHFQCN--- 144
 Db 863 QADPTNSNTGLKXKRTTLTGNISVEGNLSLTGANANTVGNLSTABDSTFKGEASDNLNITG 922
 QY 145 ---KNAASGASLYLIPAGELKNLPFGGIWDATPLKRVKRYSETYGT-VT-NITIKLATD 200
 Db 923 TPTNNTANINIKGVVKGIDNNK--GG-----UNITNASGTOKTLINGNIT---NE 970
 QY 201 KGNIQWLPQKSDARVDL--NLRP7TGGGTYIGRNSVDMCFYDGYSTNSSLEIRFQ--- 255
 Db 971 KGLDNI--KNIKADAIQIGENISQKEGNLTSSDKVNI-----TNQ7TIKAGVEGGR 1021
 QY 256 -DNKPSDGGKYRKLNDTTEKATYILSLLAG---KSLTPTNGTSLNIADASLETNWN 311
 Db 1022 SDSSEANLTIQ-----TKELKAGDLNINISGNFKAEITAKNGSDLTIGNAGGNADAX 1076
 QY 312 RITAVTMPETISVPVLWPGF-RLQIDAKVENPEAGQYMGINVT 353
 Db 1077 K---VTPDRKDSKISTDGINVTINSEKVTSSGNNAGNDNST 1116

RESULT 10

US-10-369-493-22177
 ; Sequence 22177, Application US/10369493
 ; Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ; FILE REFERENCE: 38-10(52352)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 22177
 ; LENGTH: 650
 ; TYPE: PRT
 ; ORGANISM: Saccharomyces cerevisiae
 US-10-369-493-22177

Query Match 6.0%; Score 114; DB 15; Length 650;

Best Local Similarity 20.9%; Pred. No. 0.041; Mismatches 147; Indels 172; Gaps 23;
 Matches 98; Conservative 52;

QY 4 ILFTFTLFFSVSL-----VSADKIPG-----DESITNIPGRDRNE--- 43
 Db 8 ILNLFSLALACAININDTFSNLEITPLTANKQPQGWTAITPDSIADASSIREGDEFTL 67
 QY 44 SSP---KHNLNNHITA-----YSESHLYDRMTPLC-----LSSHNTLN 80
 Db 63 SMPHVYRIKLLNSQTATISLADGTEAFKCYVSOQAAYLYENTFTCTAQNDLSSYNTID 127
 QY 81 GACPTSENPSSSVSGETNITLQFTKESLIK33-QIKGVKQLLFKSNVC-FSGLTLNSA 139
 Db 123 GSITFSLNPFSGGSGSYVEL-----ENAKFKSGPMVLKLGNDVDVNFDAFTENVF 182
 QY 143 HENCNKNAASGASLYL--YIPAGELKNLPFGGINDATLKLVRKRYSETYGYTNYTINIK 197
 Db 183 HSGRSTGYGSPSYHLGMVCPNGY?---LGG-----TEKI-----DYDSSNNVDL- 225
 QY 193 LFDKGMQ?-----WLPQKSDARVDLNLRTGGTYIGRN---SVMCFYDG--- 242
 Db 225 --DCSSVQVYSSNDENDWFFQSYNDTNADV-----TCFGSNLWITLDEKLYDGEYL 275
 QY 243 ---YSTNSSSLE--IRFQDNK--PKSDGKF 265
 Db 275 WYNALQSLPANVNTIDHALEFYQYCLDTTANTYATQFSTTREFIVYQGRNLGTASAKSS 335
 QY 265 YLRKINDPKEI---AYTLS-----LLLAGKSLTPTNGTSLNIADAA 304
 Db 336 PSTYTTDLTSLNTSAYSTGSIWETGNRTTSEVISHVVTSTKLSPATTSLTIAQTS 395
 QY 305 SLETNWN-----RITAVTMPETISVPVLWPGRL 332
 Db 396 IYSTDSNITWGTDIHTTSEVISDVETISREIATSTVVAAPTSTTGTGTCAM 444

RESULT 11

US-10-282-122A-61451
 ; Sequence 61451, Application US/10282122A
 ; Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.334A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 41
; LENGTH: 1005
; TYPE: PRF
; ORGANISM: Haemophilus influenzae
US-10-193-764-41

```

Query Match	5.9%;	Score 111.5;	DB 14;	Length 1305;
Best Local Similarity	20.3%;	Pred. No. 0.14;		
Matches	58;	Conservative 65;	Mismatches 147;	Indels 55; Gaps 14;
QY	60	ESHTLYDRMTFLCLSSHNTLACPFSENPSSSVSGETNITLQTFEKS-LIKRELQIK 118		
Db	363	EKVAIFSTNLTILGNVTILGGNSSNITKGNININSKANVTLOAHGSHLDELKE- 418		
QY	119	GYKQLLFKSWNCPISGLTL--NSAHFNCNKNAASGASLYIYPAGEL-KNLPFGGIW--DA 173		
Db	419	--RLTILGNVSGCNLIITCSNAHIDGSLIAESAKP-----QGKTNNNLITGTETNNG 471		
QY	174	TLKLRVKRYSEFYGYVTNITITIKLTD-----KGNI-----QIWLPPQKSDARVD 218		
Db	472	TADINIKQGVVLQGGITNNGNINITTRASVNOKITINGNITNKGDLNFKDIFAKAEIQ 531		
QY	219	L--NLAPTGGTIVIGNSVDMCYDGVSTKSSLETRFQDNPKSD-GKFYLKKNIDDDTK 275		
Db	532	IGGNIHQEGNLTISDKINI-----TKRIEIKADTDQGNSDSGVASNANLTIKTK 582		
QY	276	EIAVTLSSLLAG--KSLTPTNGTSLNIADAASLETNNRITAVTMEIISVPVLCWPGRLL 332		
Db	593	ELTLTUNLITSGNKAIEITAKDNLSDLLIGKASSDNSNAKOITFDKVKDSKIS--AGNHNV 640		
QY	333	QLDAKVENPEA-----GQYNGNINVTFTSSQTL 361		
Db	641	TTNSKVFSTNSDGTGNGSDNNIGHTISAKDVTV 675		

```

RESULT 13
US-10-193-764-39
; Sequence 39, Application US/10193764
; Publication No. US2003013943A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
; TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
; FILE REFERENCE: 1038-1239MIS
; CURRENT APPLICATION NUMBER: US/10/193,764
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 09/167,568
; PRIOR FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 39
; LENGTH: 1011
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-193-764-39

```

```

Query Match          5.9%; Score 111.5; DB 14; Length 1011;
Best Local Similarity 20.3%; Pred. No. 0.14;
Matches 68; Conservative 65; Mismatches 147; Indels 55; Gaps 14

50 BSHTYDRWTFCLCSHTLNGACPTSNPSSSVSGTNTLQFTKES-LIKELOIK 118
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
359 EKNAFSTHNTILGGVNTLGGSSSSNIKGININSKANVTLOAHGTSHLDKKE--- 424
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

119 GYKQLLFKSVNCPGSLT--NSAHNCKNKAASGLKLYTPAGEL-KNLPFGGIW--DA 173
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

425 --PTILGNVSGVGNLIIGSAHIDGNLSIAESAKF-----QGTNNNNLITGFTNNG 477
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

174 TLKLAVKRYSTGYCTYTNITIKLTD-----KGNL-----QIWLQFKSDARVD 218
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/2306,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 61451
/ LENGTH: 513
/ TYPE: PRT
/ ORGANISM: Legionella pneumophila
/ JS-10-282-122A-61451

```

```

Query Match      6.08; Score 113.5; DB 12; Length 513;
Best Local Similarity 21.7%; Pred. No. 0.032;
Matches          76; Conservative 51; Mismatches 116; Indels 107; Gaps 16;

QY   50 ILNHITAYSESHTLYDRMTF---LCISSHNTLNGACPTSENFSSSVSGTNTITLQFT 105
    :|::||::||::||::||::||::||::||::||::||::||::||::||::||
Db   90 LLMKLTPLLKGHVFSBEINIDGLKLPIODN-----NEKLSSEKANQVNTITYP 141
    :|::||::||::||::||::||::||::||::||::||::||::||::||

QY   106 EKSLIKGE-QIKGYKOLLFKSVCPSGLTNSAHFNCKNAASGASLYIYPAGEIKNL 165
    :|::||::||::||::||::||::||::||::||::||::||::||::||
Db   142 NEQPAIDAKLLLSHG--QLILNS-----KGHSIVLVKNIQIGAEQF-----NLKNS 183
    :|::||::||::||::||::||::||::||::||::||::||::||::||

QY   166 PFGGIWDATLKRVRKRYSETGYGTINTITIKLTDKGNIIWLQFSDARVDLNLPFG 225
    :|::||::||::||::||::||::||::||::||::||::||::||::||
Db   184 PF-----SVQIKAKLTDA-----FLQTAKANINFK----- 209
    :|::||::||::||::||::||::||::||::||::||::||::||::||

QY   226 GGTVGRRNVDMCYDGVST--NESSHIREQDNPKSCOGKYLELKINDTK-----EIAY 279
    :|::||::||::||::||::||::||::||::||::||::||::||::||
Db   210 ----GRVSLSPSIIDEINSGISKSIIEGOLQ-QNILN-QFAIKKINTTTKHKRDQIF 263
    :|::||::||::||::||::||::||::||::||::||::||::||::||

QY   280 ---TSL------LLACKSLTPNG----TSLNIADAASLETWNRITA 315
    :|::||::||::||::||::||::||::||::||::||::||::||::||
Db   284 NPLTSLNYNGESIGDMGVVIATQQULINQTAIVNDKQILTSLKHPPAISGLNDYSIHAS 323
    :|::||::||::||::||::||::||::||::||::||::||::||::||

QY   316 VTMPEISVPVLWCPRLOJ--DAKVENPEAGOVGMGINV-----TFTDSS 358
    :|::||::||::||::||::||::||::||::||::||::||::||::||
Db   324 TDKVMSIESVSZCTITKGEVYNINIDOIINNKKVLSLMTETEDN 373
    :|::||::||::||::||::||::||::||::||::||::||::||::||

```

```

RESULT 12
US-10-193-764-41
; Sequence 41, Application US/10193764
; Publication No. US20030133943A1
; GENERAL INFORMATION:
; APPLICANT: Loomsore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel R.
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
; TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
; FILE REFERENCE: 1038-1239MIS
; CURRENT APPLICATION NUMBER: US/10/193,764
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 09/167,569
; PRIOR FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 91

```

478 TADINIKOGVVKLQGDITNGNHNITNASVNQKTIINGNITNKGLDLNLIKDKANAEIQ 537
219 L--NLPTGGTYIGRNSVDMCFYDGYSTNSSSLEIRFQDNNPKSD-GKPYLRKINDDTK 275
538 CGNISQKEGNTISSDKINI-----TKRIBKADTDQNSDSGVASNANLTIKTK 588
276 EYATYLSLLAG---KSLTPTNGTSLINIDASLETNWNRITAVTWPESVPLVCPGRL 332
589 ELTLDNLNISGFNAEITAKONSDDLIIGKASDKNAXQITFDKVKCKSKIS--AGNENV 646
333 QLDKAVENPEA-----GQYMGINKVTFPPSSQTL 361
647 TUNSKVETSNSGSGTNGSGDDNNGIGTISAKOVTV 681

RESULT 14
US-09-912-020-302
; Sequence 302, Application US/09912020
; Patent No. US20020045592A1
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ESCHERICHIA COLI
; CURRENT APPLICATION NUMBER: US/09/912,020
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 302
; LENGTH: 2383
; TYPE: PRF
; ORGANISM: E. Coli
US-09-912-020-302

Query Match 5.8%; Score 109; DB 9; Length 2383;
Best Local Similarity 20.0%; Pred. No. 0.97;
Matches 78; Conservative 65; Mismatches 157; Indels 90; Gaps 15;

13 SSVLFTFAVADKIPGDESITNIGPRDRNESSPKHILNNHITAY---SESHTLYDRMT 69
762 SAKIATLSASNGVLANENAAVTYSVNVADSGS---NPINDHTVTFVAVLSGSAISFNQ 813
70 FCLSSHTNLGACPTSENPSSSVSGSTNITLOFTKRSII-----K 112
819 ----TAKTDVNGLA-TFPLKSKQEDNTVETVLENGVKQTLIVSFVGSSTAQVDLQSK 873
113 RELQIKGYKQL-LFKSVNCPGSLTNSAHFNCKNNAAGSGLYLYIPAGELKNIPLFGGIW 171
874 NEVVADGNDSTVMTATVRDAGKNLNDVWVTFNVNSAAKLSQTEVNSHD-----GIA 926
172 DATLKLVRKRYSETYGTYYTINIKLTDKNIQIWLQPKSDARVDNLNRPSTGGTVIG 231
927 TATLT-----SLKNGDYRTASVSSGSQAQNVNFTGDQSTAAITLSV-PSGDIIV-- 976
232 RNSVDMCFYDGYSTNSG---SLEIRFQDNNPKSDGKPYLRKINDDTKEIATYLSLLAG 287
977 -----TMTAQYMTATLQDKNGNPLKDKELTFSPND-----VASKFSISNGG 1019
288 KSLTPTNGTSL-----NIADAASLETNWNRITAVTWPESVPLVCPGRL 327
1020 KGMTDSNGVAIASLTGLTAGTHIMVARLANSVNSDAQPMFTVADKDRVAVVLIQTSKABII 1079

328 WPG--RLQLDAKVENPEAGQYMGINVTFT 355
1080 GNGVDETLTATVKDP-SNHPVAGITVNT 1108

RESULT 15
US-10-282-122A-42852
; Sequence 42852, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42852
; LENGTH: 2383
; TYPE: PRF
; ORGANISM: Escherichia coli
US-10-282-122A-42852

Query Match 5.8%; Score 109; DB 12; Length 2383;
Best Local Similarity 20.0%; Pred. No. 0.97;
Matches 78; Conservative 65; Mismatches 157; Indels 90; Gaps 15;

13 SSVLFTFAVADKIPGDESITNIGPRDRNESSPKHILNNHITAY---SESHTLYDRMT 69
762 SAKIATLSASNGVLANENAAVTYSVNVADSGS---NPINDHTVTFVAVLSGSAISFNQ 818
70 FCLSSHTNLGACPTSENPSSSVSGSTNITLOFTKRSII-----K 112
819 ----TAKTDVNGLA-TFPLKSKQEDNTVETVLENGVKQTLIVSFVGSSTAQVDLQSK 873
113 RELQIKGYKQL-LFKSVNCPGSLTNSAHFNCKNNAAGSGLYLYIPAGELKNIPLFGGIW 171
874 NEVVADGNDSTVMTATVRDAGKNLNDVWVTFNVNSAAKLSQTEVNSHD-----GIA 926
172 DATLKLVRKRYSETYGTYYTINIKLTDKNIQIWLQPKSDARVDNLNRPSTGGTVIG 231

Db 927 TATLT-----SLKNGDYRVTA SVSSGQANQQVNF IGDSQSTAALTLSV-PSGDITV-- 976
QY 232 RNSVDXCFYDGYSTNS- ---SLEIRFQDNPKSDGKFYLRKINDDTKEIAYTSLILLAG 287
Db 977 -----TNTAPQYMTATLQDRKNGNPLKDKKEITFSVPND-----VASKFSISNGG 1019
QY 288 KSLTPTNGTSL-----NIADAASLSTNNNRITAVTMTPEISVPVLC 327
Db 1020 KGMTDSNGVAIASLTGTIAGTHIMARLANGSNVSDAQPMTFVADKDRVVVLQTSKAEII 1079
QY 328 WPG--RLQLDKVENPEACQVMGNINVTFT 355
Db 1080 GNGVDETTLTATVKDP-SNHPVAGITVNET 1108

Search completed: May 6, 2004, 10:25:06
Job time : 49 secs